B13

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/10, 15/63, 1/21	A1	(11) International Publication Number: WO 93/08278
		(43) International Publication Date: 29 April 1993 (29.04.93)

(21) International Application Number: PCT/US92/08879

(22) International Filing Date: 15 October 1992 (15.10.92)

(30) Priority data:
07/778,233   16 October 1991 (16.10.91)   US

(60) Parent Application or Grant
  (63) Related by Continuation
    US   07/778,233 (CIP)
    Filed on   16 October 1991 (16.10.91)

(71) Applicant (for all designated States except US): AFFYMAX TECHNOLOGIES N.V. [NL/NL]; De Ruyderkade 62, Curaçao (AN).

(72) Inventors; and
(75) Inventors/Applicants (for US only) : SCHATZ, Peter, J. [US/US]; 438 Gabilan Street, #4, Los Altos, CA 94022 (US). CULL, Millard, G. [US/US]; 1307 Trestle Glen Road, Oakland, CA 94610 (US). MILLER, Jeff, F. [US/US]; 3526 Purdue Avenue, Los Angeles, CA 90066 (US). STEMMER, Willem, Peter, Christian [US/US]; 1012 Windson Drive, Menlo Park, CA 94025 (US).

(74) Agent: SMITH, William, M.; Townsend and Townsend, One Market Plaza, 20th Fl., Steuart Tower, San Francisco, CA 94105 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
  With international search report.

(54) Title: PEPTIDE LIBRARY AND SCREENING METHOD

(57) Abstract

A random peptide library constructed by transforming host cells with a collection of recombinant vectors that encode a fusion protein comprised of a DNA binding protein and a random peptide and also contain a binding site for the DNA binding protein can be used to screen for novel ligands. The screening method results in the formation of a complex comprising the fusion protein bound to a receptor through the random peptide ligand and to the recombinant DNA vector through the DNA binding protein.

<u>PEPTIDE LIBRARY AND SCREENING METHOD</u>

FIELD OF THE INVENTION

The present invention relates generally to methods for selecting peptide ligands to receptor molecules of interest and, more particularly, to methods for generating and screening large peptide libraries for peptides with desired binding characteristics.

BACKGROUND OF THE INVENTION

The isolation of ligands that bind biological receptors is fundamental to understanding signal transduction and to discovering new therapeutics. The ability to synthesize DNA chemically has made possible the construction of extremely large collections of nucleic acid and peptide sequences as potential ligands. Recently developed methods allow efficient screening of libraries for desired binding activities (see Pluckthun and Ge, 1991, <u>Angew. Chem. Int. Ed. Engl</u>. <u>30</u>:296-298). For example, RNA molecules with the ability to bind a particular protein (see Tuerk and Gold, 1990, <u>Science</u> <u>249</u>:505-510) or a dye (see Ellington and Szostak, 1990, <u>Nature</u> <u>346</u>:818-822) have been selected by alternate rounds of affinity selection and PCR amplification. A similar technique was used to determine the DNA sequences that bound a human transcription factor (see Thiesen and Bach, 1990, <u>Nucl. Acids Res</u>. <u>18</u>:3203-3209).

Application of efficient screening techniques to peptides requires the establishment of a physical or logical connection between each peptide and the nucleic acid that encodes the peptide. After rounds of affinity enrichment, such a connection allows identification, usually by amplification and sequencing, of the genetic material encoding interesting peptides. Several phage based systems for screening proteins and polypeptides have been described. The fusion phage approach of Parmley and Smith, 1988, <u>Gene</u> <u>73</u>:305-318, can be used to screen proteins. Others have described phage based systems in which the

peptide is fused to the pIII coat protein of filamentous phage
(see Scott and Smith, 1990, Science 249:386-390; Devlin et al.,
1990, Science 249:404-406; and Cwirla et al., 1990, Proc. Natl.
Acad. Sci. USA 87:6378-6382; each of which is incorporated herein
5   by reference).

In these latter publications, the authors describe
expression of a peptide at the amino terminus of or internal to
the pIII protein. The connection between peptide and the genetic
material that encodes the peptide is established, because the
10   fusion protein is part of the capsid enclosing the phage genomic
DNA. Phage encoding peptide ligands for receptors of interest
can be isolated from libraries of greater than 10^8 peptides after
several rounds of affinity enrichment followed by phage growth.
Other non-phage based systems that could be suggested for the
15   construction of peptide libraries include direct screening of
nascent peptides on polysomes (see Tuerk and Gold, supra) and
display of peptides directly on the surface of E. coli. As in
the filamentous phage system, all of these methods rely on a
physical association of the peptide with the nucleic acid that
20   encodes the peptide.

There remains a need for methods of constructing
peptide libraries in addition to the methods described above.
For instance, the above methods do not provide random peptides
with a free carboxy terminus, yet such peptides would add
25   diversity to the peptide structures now available for receptor
binding. In addition, prior art methods for constructing random
peptide libraries cannot tolerate stop codons in the degenerate
region coding for the random peptide, yet stop codons occur with
frequency in degenerate oligonucleotides. Prior art methods
30   involving phage fusions require that the fusion peptide be
exported to the periplasm and so are limited to fusion proteins
that are compatible with the protein export apparatus and the
formation of an intact phage coat.

The present invention provides random peptide libraries
35   and methods for generating and screening those libraries with
significant advantages over the prior art methods.

SUMMARY OF THE INVENTION

The present invention provides random peptide libraries and methods for generating and screening those libraries to identify peptides that bind to receptor molecules of interest. The peptides can be used for therapeutic, diagnostic, and related purposes, e.g., to bind the receptor or an analogue of the receptor and so inhibit or promote the activity of the receptor.

The peptide library of the invention is constructed so that the peptide is expressed as a fusion product; the peptide is fused to a DNA binding protein. The peptide library is constructed so that the DNA binding protein can bind to the recombinant DNA expression vector that encodes the fusion product that contains the peptide of interest. The method of generating the peptide library of the invention comprises the steps of (a) constructing a recombinant DNA vector that encodes a DNA binding protein and contains a binding site for the DNA binding protein; (b) inserting into the coding sequence of the DNA binding protein in the vector of step (a) a coding sequence for a peptide such that the resulting vector encodes a fusion protein composed of the DNA binding protein and the peptide; (c) transforming a host cell with the vector of step (b); and (d) culturing the host cell transformed in step (c) under conditions suitable for expression of the fusion protein.

The screening method of the invention comprises the steps of (a) lysing the cells transformed with the peptide library under conditions such that the fusion protein remains bound to the vector that encodes the fusion protein; (b) contacting the fusion proteins of the peptide library with a receptor under conditions conducive to specific peptide - receptor binding; and (c) isolating the vector that encodes a peptide that binds to said receptor. By repetition of the affinity selection process one or more times, the plasmids encoding the peptides of interest can be enriched. By increased stringency of the selection, peptides of increasingly higher affinity can be identified.

The present invention also relates to recombinant DNA vectors useful for constructing the random peptide library, the random peptide library, host cells transformed with the

recombinant vectors of the library, and fusion proteins expressed by those host cells.

BRIEF DESCRIPTION OF THE DRAWINGS

5   Figure 1 shows a recombinant vector of a random peptide library of the invention. In this embodiment of the invention, the DNA binding protein is the lacI gene product, the fusion protein forms a tetramer, and the tetramer interacts with the vector and immobilized receptor, as shown in the Figure. The

10   library plasmid carries the lacI gene with random coding sequence fused to the 3' end of the coding sequence of the gene, as well as two lacO sequences. The lac repressor-peptide fusions produced by the hybrid genes bind to the lacO sites on the same plasmid that encodes them. After lysis of cells containing the

15   random library, those plasmid-repressor-peptide complexes that specifically bind a chosen receptor are enriched by avidity panning against the immobilized receptor. Transformation of E. coli with recovered plasmids allows additional rounds of panning or sequencing of isolated clones.

20   Figure 2 (SEQ. ID NOS.:1-6) shows a partial restriction site, DNA sequence, and function map of plasmid pMC5. Hybridization of oligonucleotide ON-332 to oligonucleotides ON-369 and ON-370 produces a fragment with cohesive ends compatible with SfiI, HindIII digested plasmid pMC5. The

25   ligation product adds sequence coding for twelve random amino acids to the end of lacI through a six codon linker. The library plasmid also contains: the rrnB transcriptional terminator, the bla gene to permit selection on ampicillin, the M13 phage intragenic region to permit rescue of single-stranded DNA, a

30   plasmid replication origin (ori), two lacO$_S$ sequences, and the araC gene to permit positive and negative regulation of the araB promoter that drives expression of the lacI fusion gene.

       Figure 3 (SEQ. ID NOS.:7-64) shows sequences isolated by panning with the D32.39 antibody. Each sequence is listed

35   with a clone number, the panning round in which the clone was isolated, and the result of the ELISA with D32.39 antibody. The sequences are aligned to show the D32.39 epitope that they share (box).

Figure 4 shows the linker sequences from vectors pJS141 and pJS142.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

5   For purposes of clarity and a complete understanding of the invention, the following terms are defined.

"DNA Binding Protein" refers to a protein that specifically interacts with deoxyribonucleotide strands. Those of skill in the art will recognize that, for purposes of the present invention, the DNA binding protein must bind specifically to a recombinant DNA vector and, in a preferred embodiment, bind to a specific sequence of DNA contained in the vector. In embodiments of the invention in which RNA vectors are used, DNA binding protein can also refer to an RNA binding protein.

15   "Epitope" refers to that portion of an antigen that interacts with an antibody.

"Host Cell" refers to a eukaryotic or procaryotic cell or group of cells that can be or has been transformed by a recombinant DNA vector. For purposes of the present invention, a host cell is typically a bacterium, such as an E. coli K12 cell or an E. coli B cell.

"Ligand" refers to a molecule, such as a random peptide, that is recognized by a particular receptor.

"Ligand Fragment" refers to a portion of a gene encoding a ligand and to the portion of the ligand encoded by that gene fragment.

"Ligand Fragment Library" refers not only to a set of recombinant DNA vectors that encodes a set of ligand fragments, but also to the set of ligand fragments encoded by those vectors, as well as the fusion proteins containing those ligand fragments.

"Linker" or "spacer" refers to a molecule or group of molecules that connects two molecules, such as a DNA binding protein and a random peptide, and serves to place the two molecules in a preferred configuration, e.g., so that the random peptide can bind to a receptor with minimal steric hindrance from the DNA binding protein.

"Peptide" or "polypeptide" refers to a polymer in which the monomers are alpha amino acids joined together through amide

bonds. Peptides are two or often more amino acid monomers long. Standard abbreviations for amino acids are used herein (see Stryer, 1988, _Biochemistry_, Third Ed., incorporated herein by reference.)

5   "Random Peptide" refers to an oligomer composed of two or more amino acid monomers and constructed by a stochastic or random process. A random peptide can include framework or scaffolding motifs, as described below.

           "Random Peptide Library" refers not only to a set of
10 recombinant DNA vectors that encodes a set of random peptides, but also to the set of random peptides encoded by those vectors, as well as the fusion proteins containing those random peptides.

           "Receptor" refers to a molecule that has an affinity for a given ligand. Receptors can be naturally occurring or
15 synthetic molecules. Receptors can be employed in an unaltered state or as aggregates with other species. Receptors can be attached, covalently or noncovalently, to a binding member, either directly or via a specific binding substance. Examples of receptors include, but are not limited to, antibodies, including
20 monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells, or other materials), cell membrane receptors, enzymes, and hormone receptors.

           "Recombinant DNA Vector" refers to a DNA or RNA
25 molecule that encodes a useful function and can be used to transform a host cell. For purposes of the present invention, a recombinant DNA vector typically is a phage or plasmid and can be extrachromosomally maintained in a host cell or controllably integrated into and excised from a host cell chromosome.

30   The present invention provides random peptide libraries and methods for generating and screening those libraries to identify either peptides that bind to receptor molecules of interest or gene products that modify peptides or RNA in a desired fashion. The peptides are produced from libraries of
35 random peptide expression vectors that encode peptides attached to a DNA binding protein. A method of affinity enrichment allows a very large library of peptides to be screened and the vector carrying the desired peptide(s) to be selected. The nucleic acid

can then be isolated from the vector and sequenced to deduce the amino acid sequence of the desired peptide. Using these methods, one can identify a peptide as having a desired binding affinity for a molecule. The peptide can then be synthesized in bulk by
5   conventional means.

By identifying the peptide de novo, one need not know the sequence or structure of the receptor molecule or the sequence or structure of the natural binding partner of the receptor. Indeed, for many "receptor" molecules a binding
10   partner has not yet been identified. A significant advantage of the present invention is that no prior information regarding an expected ligand structure is required to isolate peptide ligands of interest. The peptide identified will have biological activity, which is meant to include at least specific binding
15   affinity for a selected receptor molecule and, in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like.

20   The number of possible receptor molecules for which peptide ligands may be identified by means of the present invention is virtually unlimited. For example, the receptor molecule may be an antibody (or a binding portion thereof). The antigen to which the antibody binds may be known and perhaps even
25   sequenced, in which case the invention may be used to map epitopes of the antigen. If the antigen is unknown, such as with certain autoimmune diseases, for example, sera, fluids, tissue, or cell from patients with the disease can be used in the present screening method to identify peptides, and consequently the
30   antigen, that elicits the autoimmune response. One can also use the present screening method to tailor a peptide to a particular purpose. Once a peptide has been identified, that peptide can serve as, or provide the basis for, the development of a vaccine, a therapeutic agent, a diagnostic reagent, etc.

35   The present invention can be used to identify peptide ligands for a wide variety of receptors in addition to antibodies. These ligands include, by way of example and not limitation, growth factors, hormones, enzyme substrates,

interferons, interleukins, intracellular and intercellular messengers, lectins, cellular adhesion molecules, and the like. Peptide ligands can also be identified by the present invention for molecules that are not peptides or proteins, e.g.,

5 carbohydrates, non-protein organic compounds, metals, etc. Thus, although antibodies are widely available and conveniently manipulated, antibodies are merely representative of receptor molecules for which peptide ligands can be identified by means of the present invention.

10   The peptide library is constructed so that the DNA binding protein-random peptide fusion product can bind to the recombinant DNA expression vector that encodes the fusion product that contains the peptide of interest. The method of generating the peptide library comprises the steps of (a) constructing a

15 recombinant DNA vector that encodes a DNA binding protein and contains binding sites for the DNA binding protein; (b) inserting into the coding sequence of the DNA binding protein in a multiplicity of vectors of step (a) coding sequences for random peptides such that the resulting vectors encode different fusion

20 proteins, each of which is composed of the DNA binding protein and a random peptide; (c) transforming host cells with the vectors of step (b); and (d) culturing the host cells transformed in step (c) under conditions suitable for expression of the fusion proteins. Typically, a random peptide library will

25 contain at least 10^6 to 10^8 different members, although library sizes of 10^8 to 10^{13} can be achieved.

     The peptide library produced by this method is especially useful in screening for ligands that bind to a receptor of interest. This screening method comprises the steps

30 of (a) lysing the cells transformed with the peptide library under conditions such that the fusion protein remains bound to the vector that encodes the fusion protein; (b) contacting the fusion proteins of the peptide library with a receptor under conditions conducive to specific peptide – receptor binding; and

35 (c) isolating the vector that encodes a peptide that binds to said receptor. By repetition of the affinity selection process one or more times, the vectors that encode the peptides of interest may be enriched. By increased stringency of the

selection, peptides of increasingly higher affinity can be
identified.   If the presence of cytoplasmic or periplasmic
proteins interferes with binding of fusion protein to receptor,
then partial purification of fusion protein-plasmid complexes by
5 gel filtration, affinity, or other purification methods can be
used to prevent such interference. For instance, purification of
the cell lysate on a column (such as the Sephacryl S-400 HR
column) that removes small proteins and other molecules may be
useful.

10   The recombinant vectors of the random peptide library
are constructed so that the random peptide is expressed as a
fusion product; the peptide is fused to a DNA binding protein.
A DNA binding protein of the invention must exhibit high avidity
binding to DNA and have a region that can accept insertions of
15 amino acids without interfering with the DNA binding activity.
The half-life of a DNA binding protein-DNA complex produced by
practice of the present method must be long enough to allow
screening to occur. Typically, the half-life will be one to four
hours or longer.

20   Suitable DNA binding proteins for purposes of the
present invention include proteins selected from a large group of
known DNA binding proteins including transcriptional regulators
and proteins that serve structural functions on DNA. Examples
include: proteins that recognize DNA by virtue of a
25 helix-turn-helix motif, such as the phage 434 repressor, the
lambda phage cI and cro repressors, and the E. coli CAP protein
from bacteria and proteins from eukaryotic cells that contain a
homeobox helix-turn-helix motif; proteins containing the
helix-loop-helix structure, such as myc and related proteins;
30 proteins with leucine zippers and DNA binding basic domains such
as fos and jun;   proteins with 'POU' domains such as the
Drosophila paired protein;   proteins with domains whose
structures depend on metal ion chelation such as Cys_2His_2 zinc
fingers found in TFIIIA, $Zn_2(Cys)_6$ clusters such as those found
35 in yeast Gal4, the Cys_3His box found in retroviral nucleocapsid
proteins, and the $Zn_2(Cys)_8$ clusters found in nuclear hormone
receptor-type proteins; the phage P22 Arc and Mnt repressors
(see Knight et al., 1989, J. Biol. Chem. 264(7):3639-3642 and

Bowie and Sauer, 1989, J. Biol. Chem. 264(13):7596-7602, each of
which is incorporated herein by reference); and others. Proteins
that bind DNA in a non-sequence-specific manner could also be
used, for example, histones, protamines, and HMG type proteins.

5   In addition, proteins could be used that bind to DNA indirectly,
by virtue of binding another protein bound to DNA. Examples of
these include yeast Gal80 and adenovirus E1A protein.

Although many DNA binding proteins can be used in the
construction of the libraries of the invention, the lac repressor

10   is preferred for several reasons. The lac repressor, a 37 kDa
protein, is the product of the E. coli lacI gene and negatively
controls transcription of the lacZYA operon by binding to a
specific DNA sequence called lacO. Structure-function
relationships in the lac repressor have been studied extensively

15   through the construction of thousands of amino acid substitution
variants of the protein (see Gordon et al., 1988, J. Mol. Biol.
200:239-251, and Kleina and Miller, 1990, J. Mol. Biol.
212:295-318). The repressor exists as a tetramer in its native
form with two high affinity DNA binding domains formed by the

20   amino termini of the subunits (see Beyreuther, 1980, The Operon
(Miller and Reznikoff, eds., Cold Spring Harbor Laboratory), pp.
123-154). The two DNA binding sites exhibit strong cooperativity
of binding to DNA molecules with two lacO sequences. A single
tetramer can bind to suitably spaced sites on a plasmid, forming

25   a loop of DNA between the two sites, and the resulting complex is
stable for days (see Besse et al., 1986, EMBO J. 5:1377-1381;
Flashner and Gralla, 1988, Proc. Natl. Acad. Sci. USA
85:8968-8972; Hsieh et al., 1987, J. Biol. Chem. 262:14583-14591;
Kramer et al., 1987, EMBO J. 6:1481-1491; Mossing and Record,

30   1986, Science 233:889-892; and Whitson et al., 1987, J. Biol.
Chem. 262:14592-14599).

The carboxy terminal domains of the lac repressor form
the dimer and tetramer contacts, but significantly, fusions of
proteins as large as β-galactosidase can be made to the carboxy

35   terminus without eliminating the DNA binding activity of the
repressor (see Muller-Hill and Kania, 1974, Nature 249:561-563;
and Brake et al., 1978, Proc. Natl. Acad. Sci. USA 75:4824-4827).
The lac repressor fusion proteins of the present invention

include not only carboxy terminus fusions but also amino terminus fusions and peptide insertions in the lac repressor. Substitutions of other sequences, including eukaryotic nuclear localization signals, transcriptional activation domains, and

5   nuclease domains, have been made at both the amino and carboxy termini of the lac repressor without serious disruption of specific DNA binding (see Hu and Davidson, 1991, Gene 99:141-150; Labow et al., 1990, Mol. Cell. Biol. 10:3343-3356; and Panayotatos et al., 1989, J. Biol. Chem. 264:15066-15069).

10   The binding of the lac repressor to a single wild-type lacO is both tight and rapid, with a dissociation constant of 10^{-13} M, an association rate constant of 7 x 10^9 $M^{-1}s^{-1}$, and a half-life for the lac repressor-lacO complex of about 30 min. (see Barkley and Bourgeois, 1980, The Operon (Miller and

15   Reznikoff, eds., Cold Spring Harbor Laboratory), pp. 177-220). The high stability of the lac repressor-DNA complex has permitted its use in methods for identifying DNA binding proteins (see Levens and Howley, 1985, Mol. Cell. Biol. 5:2307-2315), for quantifying PCR-amplified DNA (see Lundeberg et al., 1991,

20   Bio/Tech. 10:68-75), and for cleavage of the E. coli and yeast genomes at a single site (see Koob and Szybalski, 1990, Science 250:271-273). This stability is important for purposes of the present invention, because, for the affinity selection or "panning" step of the screening process to succeed, the

25   connection between the fusion protein and the plasmid that encodes the fusion protein must remain intact for at least a portion of the complexes throughout the panning step.

In fact, for purposes of the present invention, a longer half-life is preferred. A variety of techniques can be

30   used to increase the stability of the DNA binding protein-DNA complex. These techniques include altering the amino acid sequence of the DNA binding protein, altering the DNA sequence of the DNA binding site, increasing the number of DNA binding sites on the vector, adding compounds that increase the stability of

35   the complex (such as lactose or ONPF for the lac system), and various combinations of each of these techniques.

An illustrative random peptide library cloning vector of the invention, plasmid pMC5, demonstrates some of these

techniques.   Plasmid pMC5 has two _lacO_ sequences to take advantage of the strong cooperative interaction between a _lac_ repressor tetramer and two _lac_ repressor binding sites, and each of these sequences is the symmetric variant of the _lacO_ sequence, called _lacO_$_s$ or _lacO_$_{id}$, which has about ten fold higher affinity for repressor than the wild-type sequence (see Sadler _et al._, 1983, Proc. Natl. Acad. Sci. USA _80_:6785-6789, and Simons _et al._, 1984, Proc. Natl. Acad. Sci. USA _81_:1624-1628).   Other "tight-binding" _lac_ repressors and coding sequences for those repressors that can be used for purposes of the present invention are described in Maurizot and Grebert, 1988, FEBS Lettrs. _239_(1):105-108, incorporated herein by reference.   See also Lehming _et al._, 1987 EMBO J. _6_(10):3145-3153.

Plasmid pMC5 is shown in Figures 1 and 2, and details of the construction of the plasmid are in Example 1, below. This library plasmid contains two major functional elements in a vector that permits replication and selection in _E. coli_. The _lacI_ gene is expressed under the control of the _araB_ promoter and has a series of restriction enzyme sites at the 3' end of the gene.   Synthetic oligonucleotides cloned into these sites fuse the _lac_ repressor protein coding sequence to additional random peptide coding sequence.

Once a vector such as pMC5 is constructed, one need only clone peptide coding sequences in frame with the DNA binding protein coding sequences to obtain a random peptide library of the invention. Thus, the random peptide library of the invention is constructed by cloning an oligonucleotide that contains the random peptide coding sequence (and any spacers, framework determinants, etc., as discussed below) into a selected cloning site of a vector that encodes a DNA binding protein and binding sites for that protein.

Using known recombinant DNA techniques (see generally, Sambrooke _et al._, 1989, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, incorporated herein by reference), one can synthesize an oligonucleotide that, _inter alia_, removes unwanted restriction sites and adds desired ones, reconstructs the correct portions of any sequences that have been removed, inserts the spacer,

conserved, or framework residues, if any, and corrects the translation frame (if necessary) to produce an active fusion protein comprised of a DNA binding protein and random peptide. The central portion of the oligonucleotide will generally contain one or more random peptide coding sequences (variable region domain) and spacer or framework residues. The sequences are ultimately expressed as peptides (with or without spacer or framework residues) fused to or in the DNA binding protein.

The variable region domain of the oligonucleotide encodes a key feature of the library: the random peptide. The size of the library will vary according to the number of variable codons, and hence the size of the peptides, that are desired. Generally, the library will be at least 10^6 to 10^8 or more members, although smaller libraries may be quite useful in some circumstances. To generate the collection of oligonucleotides that forms a series of codons encoding a random collection of amino acids and that is ultimately cloned into the vector, a codon motif is used, such as $(NNK)_x$, where N may be A, C, G, or T (nominally equimolar), K is G or T (nominally equimolar), and x is typically up to about 5, 6, 7, or 8 or more, thereby producing libraries of penta-, hexa-, hepta-, and octa-peptides or more. The third position may also be G or C, designated "S". Thus, NNK or NNS (i) code for all the amino acids, (ii) code for only one stop codon, and (iii) reduce the range of codon bias from 6:1 to 3:1. There are 32 possible codons resulting from the NNK motif: 1 for each of 12 amino acids, 2 for each of 5 amino acids, 3 for each of 3 amino acids, and only one of the three stop codons. With longer peptides, the size of the library that is generated can become a constraint in the cloning process, but the larger libraries can be sampled, as described below. The expression of peptides from randomly generated mixtures of oligonucleotides in recombinant vectors is discussed in Oliphant et al., 1986, Gene 44:177-183, incorporated herein by reference.

An exemplified codon motif $(NNK)_x$ produces 32 codons, one for each of 12 amino acids, two for each of five amino acids, three for each of three amino acids and one (amber) stop codon. Although this motif produces a codon distribution as equitable as available with standard methods of oligonucleotide synthesis, it

results in a bias against peptides containing one-codon residues.
For example, a complete collection of hexacodons contains one
sequence encoding each peptide made up of only one-codon amino
acids, but contains 729 (3^6) sequences encoding each peptide with
5   three-codon amino acids.

     An alternate approach that minimizes the bias against
one-codon residues involves the synthesis of 20 activated
trinucleotides, each representing the codon for one of the 20
genetically encoded amino acids.   These trinucleotides are
10   synthesized by conventional means, removed from the support with
the base and 5-OH-protecting groups intact, and activated by the
addition of 3'-O-phosphoramidite (and phosphate protection with
beta-cyanoethyl groups) by the method used for the activation of
mononucleosides, as generally described in McBride and Caruthers,
15   1983, Tetr. Letters 22:245, which is incorporated herein by
reference.

     Degenerate "oligocodons" are prepared using these
trimers as building blocks. The trimers are mixed at the desired
molar ratios and installed in the synthesizer. The ratios will
20   usually be approximately equimolar, but may be a controlled
unequal ratio to obtain the over- to under-representation of
certain amino acids coded for by the degenerate oligonucleotide
collection.   The condensation of the trimers to form the
oligocodons is done essentially as described for conventional
25   synthesis employing activated mononucleosides as building blocks.
See generally, Atkinson and Smith, 1984, Oligonucleotide
Synthesis (M.J. Gait, ed.), pp. 35-82. This procedure generates
a population of oligonucleotides for cloning that is capable of
encoding an equal distribution (or a controlled unequal
30   distribution) of the possible peptide sequences. This approach
may be especially useful in generating longer peptide sequences,
because the range of bias produced by the $(NNK)_x$ motif increases
by three-fold with each additional amino acid residue.

     When the codon motif is $(NNK)_x$, as defined above, and
35   when x equals 8, there are 2.6×10^{10} possible octapeptides. A
library containing most of the octapeptides may be produced, but
a sampling of the octapeptides may be more conveniently
constructed by making only a subset library using about 0.1%, and

up to as much as 1%, 5%, or 10%, of the possible sequences, which
subset of recombinant vectors is then screened. As the library
size increases, smaller percentages are acceptable. If desired,
to extend the diversity of a subset library the recovered vector
5   subset may be subjected to mutagenesis and then subjected to
subsequent rounds of screening. This mutagenesis step may be
accomplished in two general ways: the variable region of the
recovered phage may be mutagenized or additional variable amino
acids may be added to the regions adjoining the initial variable
10   sequences.

The process of constructing a random peptide encoding
oligonucleotide is described in Example 2, below. In brief, a
library can be constructed in pMC5 using the half-site cloning
strategy of Cwirla et al., supra. A random dodecamer peptide
15   sequence, connected to the C-terminus of the lac repressor
through a linker peptide GADGGA (GADGA [SEQ. ID NO.:65]) would
also be an acceptable linker), can be specified by a degenerate
oligonucleotide population containing twelve codons of the form
NNK, where N is any base, and K is G or T. Transformation of E.
20   coli strain MC1061 using 4 μg of pMC5 ligated to a four fold
molar excess of annealed oligonucleotides yielded a test library
of 5.5 x 10^8 independent clones.

Once the library is constructed, host cells are
transformed with the library vectors. The successful
25   transformants are typically selected by growth in a selective
medium or under selective conditions, e.g., an appropriate
antibiotic, which, in the case of plasmid pMC5 derivatives, is
preferably ampicillin. This selection may be done on solid or in
liquid growth medium. For growth on solid medium, the cells are
30   grown at a high density ($^-10^8$ to 10^9 transformants per m^2) on a
large surface of, for example, L-agar containing the selective
antibiotic to form essentially a confluent lawn. For growth in
liquid culture, cells may be grown in L-broth (with antibiotic
selection) through about 10 or more doublings. Growth in liquid
35   culture may be more convenient because of the size of the
libraries, while growth on solid media likely provides less
chance of bias during the amplification process.

For best results with the present method, one should control the ratio of fusion proteins to vectors so that vectors are saturated with fusion proteins, without a vast excess of fusion protein. Too little fusion protein could result in
5   vectors with free binding sites that might be filled by fusion protein from other cells in the population during cell lysis, thus breaking the connection between the genetic information and the peptide ligand. Too much fusion protein could lead to titration of available receptor sites during panning by fusion
10   protein molecules not bound to plasmid. To control this ratio, one can use any of a variety of origin of replication sequences to control vector number and/or an inducible promoter, such as any of the promoters selected from the group consisting of the araB, lambda pL, (which can be either nalidixic acid or heat
15   inducible or both), trp, lac, T7, T3, and tac or trc (these latter two are trp/lac hybrids) promoters to control fusion protein number. A regulated promoter is also useful to limit the amount of time that the peptide ligands are exposed to cellular proteases. By inducing the promoter a short time before lysing
20   the cells containing a library, one can minimize the time during which proteases act.

The araB promoter normally drives expression of the enzymes of the E. coli araBAD operon, which are involved in the catabolism of L-arabinose. The araB promoter is regulated both
25   positively and negatively, depending on the presence of L-arabinose in the growth medium, by the AraC protein. This promoter can be catabolite repressed by adding glucose to the growth medium and induced by adding L-arabinose to the medium. Plasmid pMC5 encodes and can drive expression of the AraC protein
30   (see Lee, 1980, The Operon (Miller and Reznikoff, eds., Cold Spring Harbor Laboratory), pp. 389-409). The araB promoter is also regulated by the CAP protein, an activator involved in the E. coli system of catabolite repression.

The expression level of the lacI fusion gene under the
35   control of the araB promoter in plasmid pMC5 can be controlled over a very wide range through changes in the growth medium. One can construct a vector to measure expression of a fusion protein encoding gene to determine the growth conditions needed to

maintain an acceptable ratio of repressors to vectors. Plasmid
pMC3 is such a vector and can be constructed by attaching an
oligonucleotide that encodes a short peptide linker (GADGA [SEQ.
ID NO.:65]) followed by dynorphin B (YGGFLRRQFKVVT [SEQ. ID
5   NO.:66]) to the lacI gene in plasmid pMC5. Monoclonal antibody
D32.39 binds to dynorphin B, a 13 amino acid opioid peptide (see
Barrett   and   Goldstein,   1985,   Neuropeptides   6:113-120,
incorporated herein by reference). These same reagents, plasmids
pMC3 and pMC5 and receptor D32.39, provide a test receptor and
10   positive and negative controls for use in panning experiments,
described below.   Growth of E. coli transformants harboring
plasmid pMC3 in LB broth (10 g of tryptone, 5 g of NaCl, and 5 g
of yeast extract per liter) allowed detection in a Western blot
of a faint band of the expected molecular weight, while addition
15   of 0.2% glucose rendered this band undetectable. Growth in LB
plus 0.2% L-arabinose led to the production of a very heavy band
on a stained gel, representing greater than 25% of the total cell
protein.

        To prevent overproduction of the fusion protein encoded
20   by a plasmid pMC5 derivative (or any other vector of the present
invention that has an inducible promoter), one can grow the
transformants first under non-inducing conditions (to minimize
exposure of the fusion protein to cellular proteases and to
minimize exposure of the cell to the possibly deleterious effects
25   of the fusion protein) and then under "partial induction"
conditions.   For the araB promoter, partial induction can be
achieved with as little as 3.3×10^{-5}% of L-arabinose (as
demonstrated by increased repression in the assay described
below). A preferred way to achieve partial induction consists of
30   growing the cells in 0.1% glucose until about 30 min. before the
cells are harvested; then, 0.2 to 0.5% L-arabinose is added to
the culture to induce expression of the fusion protein. Other
methods to express the protein controllably are available.

        One can estimate the lacI expression level necessary to
35   fill the available binding sites in a typical plasmid pMC5
derivative by observing the behavior of strain ARI 20 (lacI⁻
lacZYA⁺) transformed with pMC3 or pMC5 (encoding only the linker
peptide GADGA [SEQ. ID NO.:65]).   Because the lacO sites in

plasmids pMC3 and pMC5 have higher affinity than those in the
lacZYA operon, the available repressor should fill the plasmid
sites first. Substantial repression of lacZYA should be observed
only if there is an excess of repressor beyond the amount needed

5   to fill the plasmid sites. As shown by color level on X-gal
indicator plates and direct assays of β-galactosidase (see
Miller, J.H., 1972, Experiments in Molecular Genetics (Cold
Spring Harbor Laboratory, Cold Spring Harbor, NY), incorporated
herein by reference), the amount of repressor produced by pMC5 is

10 sufficient to fill the lacO sites and repress greater than 200
fold lacZYA in ARI 20 host cells during growth in normal LB
medium (2.4 units compared to 500 units from ARI 20 transformed
with vector pBAD18, which has no lacI). The repressor encoded by
pMC3 was partially inactivated by the addition of the dynorphin

15 B tail, allowing about 10 fold higher expression of lacZYA (37
units). Because of the apparent excess production of repressor
under these conditions, LB is a preferred medium for expressing
similar fusion proteins of the invention.

    At some point during the growth of the transformants,

20 the fusion protein will be expressed. Because the random peptide
vector also contains DNA binding sites for the DNA binding
protein, fusion proteins will bind to the vectors that encode
them. After these complexes form, the cells containing a library
are lysed, and the complexes are partially purified away from

25 cell debris. Following cell lysis, one should avoid cross
reaction between unbound fusion proteins of one cell with
heterologous DNA molecules of another cell. The presence of high
concentrations of the DNA binding site for the DNA binding
protein will minimize this type of cross reaction. Thus, for the

30 lac system, one can synthesize a DNA duplex encoding the lacO or
a mutated lacO sequence for addition to the cell lysis solution.
The compound ONPF, as well as lactose, is known to strengthen the
binding of the lac repressor to lacO, so one can also, or
alternatively, add ONPF or lactose to the cell lysis solution to

35 minimize this type of cross reaction.

    After cell lysis, in a process called panning,
plasmid-peptide complexes that bind specifically to immobilized
receptors are separated from nonbinding complexes, which are

washed away. Bulk DNA can be included during the lysis and panning steps to compete for non-specific binding sites and to lower the background of non-receptor-specific binding to the immobilized receptor. A variety of washing procedures can be used to enrich for retention of molecules with desired affinity ranges. For affinity enrichment of desired clones, from about 10^2 to 10^6 library equivalents (a library equivalent is one of each recombinant; 10^4 equivalents of a library of 10^9 members is 10^{13} vectors), but typically 10^3 to 10^4 library equivalents, are incubated with a receptor (or portion thereof) for which a peptide ligand is desired. The receptor is in one of several forms appropriate for affinity enrichment schemes. In one example the receptor is immobilized on a surface or particle, and the library is then panned on the immobilized receptor generally according to the procedure described below.

A second example of receptor presentation is receptor attached to a recognizable ligand (which may be attached via a spacer). A specific example of such a ligand is biotin. The receptor, so modified, is incubated with the library, and binding occurs with both reactants in solution. The resulting complexes are then bound to streptavidin (or avidin) through the biotin moiety. See PCT patent publication No. 91/07087. The streptavidin may be immobilized on a surface such as a plastic plate or on particles, in which case the complexes (vector/DNA binding protein/peptide/receptor/biotin/streptavidin) are physically retained; or the streptavidin may be labelled, with a fluorophore, for example, to tag the active fusion protein for detection and/or isolation by sorting procedures, e.g., on a fluorescence-activated cell sorter.

Vectors that express peptides without the desired specificity are removed by washing. The degree and stringency of washing required will be determined for each receptor/peptide of interest. A certain degree of control can be exerted over the binding characteristics of the peptides recovered by adjusting the conditions of the binding incubation and the subsequent washing. The temperature, pH, ionic strength, divalent cation concentration, and the volume and duration of the washing will select for peptides within particular ranges of affinity for the

receptor.   Selection based on slow dissociation rate, which is usually predictive of high affinity, is the most practical route. This may be done either by continued incubation in the presence of a saturating amount of free ligand, or by increasing the

5   volume, number, and length of the washes.   In each case, the rebinding of dissociated peptide-vector is prevented, and with increasing time, peptide-vectors of higher and higher affinity are recovered.   Additional modifications of the binding and washing procedures may be applied to find peptides that bind

10   receptors under special conditions.

Although the screening method is highly specific, the procedure generally does not discriminate between peptides of modest affinity (micromolar dissociation constants) and those of high affinity (nanomolar dissociation constants or greater).   The

15   ability to select peptides with relatively low affinity may be the result of multivalent interaction between a vector/fusion protein complex and a receptor.   For instance, when the receptor is an IgG antibody, each complex may bind to more than one antibody binding site, either by a single complex binding through

20   the multiple peptides displayed to both sites of a single IgG molecule or by forming a network of complex-IgG.   Multivalent interaction produces a high avidity and tenacious adherence of the vector during washing.   Multivalent interactions can be mimicked by using a high density of immobilized monovalent

25   receptor.

To enrich for the highest affinity peptide ligands, a substantially monovalent interaction between vector and the receptor (typically immobilized on a solid phase) may be appropriate.   The screening (selection) with substantially

30   monovalent interaction can be repeated as part of additional rounds of amplification and selection of vectors.   Monovalent interactions may be achieved by employing low concentrations of receptor, such as the Fab binding fragment of an antibody molecule.

35   A strategy employing a combination of conditions favoring multivalent or monovalent interactions can be used to advantage in producing new peptide ligands for receptor molecules.   By conducting the first rounds of screening under

conditions to promote multivalent interactions, one can then use high stringency washing to reduce greatly the background of non-specifically bound vectors. This high avidity step may select a large pool of peptides with a wide range of affinities,

5   including those with relatively low affinity. Subsequent screening under conditions favoring increasingly monovalent interactions and isolation of phage based on a slow dissociation rate may then allow the identification of the highest affinity peptides.

10   The present invention also provides univalent display vectors. As noted above, the lac operator is an inverted DNA repeat, and two separate lac repressor proteins bind to the operator. However, only a portion of the lac repressor protein binds to the operator. This portion, called the "headpiece," can

15   be made as a recombinant protein (see Kaptein et al., 1990, Bioch. Pharmacol. 40:89-96). The present invention provides a recombinant DNA vector that encodes two or more fused lac headpieces and contains a site into which random peptide coding sequences can be introduced. This vector, described in Example

20   5, provides a monovalent display system, although a more preferred version of the system would comprise only one lacO sequence per vector (the vector in Example 5 has two lacO sequences). One can also construct derivatives of such vectors that encode 3, 4, 5, or more headpieces or encode mutated

25   headpieces or binding sites (for instance, by using the Hisl repressor headpiece and the 344 operator mutants described in Lehming et al., 1987, EMBO J. 6(10):3145-3153, incorporated herein by reference) to increase the affinity of the repressor for the operator.

30   In another aspect of the present invention, one can determine the relative affinity of a series of related peptide ligands by measuring the dissociation rate for a peptide of interest and the selected receptor molecule under substantially monovalent conditions. For example, one can measure the

35   dissociation of a Fab fragment from vector/fusion protein complexes immobilized on particles. This procedure avoids the necessity and inconvenience of separately determining binding

affinities for a selected peptide, which could be especially
burdensome if a large number of peptides have been selected.

After washing the receptor-fusion protein-vector
complexes to select for peptides of the desired affinity, the
5   vector DNA is then released from bound complexes by, for example,
treatment with high salt or extraction with phenol, or both. For
the lac system, one can use IPTG, a compound known to decrease
the stability of the lac repressor-lacO complex, to dissociate
the plasmid from the fusion protein. In a preferred embodiment,
10   the elution buffer includes 1 mM IPTG, 10 µg/mL of a double-
stranded oligonucleotide that encodes $lacO_S$, and 0.05 to 0.2 M
KCl. Once released from bound complexes, the plasmids are
reintroduced into E. coli by transformation. Because of the high
efficiency, the preferred method of transformation is
15   electroporation. Using this new population of transformants, one
can repeat additional cycles of panning to increase the
proportion of peptides in the population that are specific for
the receptor. The structure of the binding peptides can then be
determined by sequencing the 3' region of the lacI fusion gene.
20   As noted above, antibody D32.39 and the pMC3 complex
serves as a receptor-ligand positive control in panning
experiments to determine ability to recover plasmids based on the
sequence of the fusion peptide. Useful negative controls are
pMC5, which encodes only the linker fusion peptide (GADGA [SEQ.
25   ID NO.:65]), and pMC1, which encodes the dynorphin B peptide, but
lacks the lacO sequences carried by pMC3 and pMC5. Lysates of E.
coli strains carrying each plasmid were panned on D32.39
immobilized on polystyrene petri dishes. After washing, plasmids
were recovered from complexes bound to the plates by phenol
30   extraction, followed by transformation of E. coli.

The results with pure lysates demonstrated about 100
fold more transformants recovered from pMC3 lysates as compared
to the negative controls. The results with mixed lysates
revealed enrichment of pMC3 versus controls among the population
35   of recovered plasmids. The results with cells that were mixed
before lysis yielded similar results. These results show that
the plasmid-lacI-peptide complexes were sufficiently stable to

allow enrichment of plasmids on the basis of the peptide the plasmids encode.

The random dodecapeptide library in pMC5 described above was used in the screening method of the invention to identify vectors that encode a fusion protein that comprised a peptide that would bind to D32.39 antibody coupled to sheep antimouse antibody coated magnetic beads. The number of complexes added to the beads at each round of panning yielded the equivalent of 10^{10} to 10^{11} transformants (see Example 3). After panning, the recovered plasmids yielded transformants ranging in number from about 10^8 in early rounds to almost 10^{11} in the fourth and final round. Compared to the number of transformants from antibody panned complexes, panning against unmodified polystyrene beads produced orders of magnitude fewer transformants.

The above results demonstrate that the DNA binding activity of _lac_ repressor can act as a link between random peptides and the genetic material encoding them and so serve as the base on which to construct large peptide ligand libraries that can be efficiently screened. In the screening process, plasmid-repressor-peptide complexes are isolated by panning on immobilized receptor, the plasmids are amplified after transformation of _E_. _coli_, and the procedure is repeated to enrich for plasmids encoding peptides specific for the receptor. The repressor binds to the library plasmid with sufficient avidity to allow panning of the library on immobilized receptor without problematic levels of dissociation. This system can be used to identify a series of related peptides that bind to a monoclonal antibody whose epitope has not been characterized and to identify peptide ligands for other receptors.

Once a peptide ligand of interest has been identified, a variety of techniques can be used to diversify a peptide library to construct ligands with improved properties. In one approach, the positive vectors (those identified in an early round of panning) are sequenced to determine the identity of the active peptides. Oligonucleotides are then synthesized based on these peptide sequences, employing all bases at each step at concentrations designed to produce slight variations of the primary oligonucleotide sequences. This mixture of (slightly)

degenerate oligonucleotides is then cloned into the random peptide library expression vector as described herein. This method produces systematic, controlled variations of the starting peptide sequences but requires, however, that individual positive

5   vectors be sequenced before mutagenesis. This method is useful for expanding the diversity of small numbers of recovered vectors.

Another technique for diversifying a selected peptide involves the subtle misincorporation of nucleotide changes in the

10   coding sequence for the peptide through the use of the polymerase chain reaction (PCR) under low fidelity conditions. A protocol described in Leung et al., 1989, Technique 1:11-15, utilizes altered ratios of nucleotides and the addition of manganese ions to produce a 2% mutation frequency.

15   Yet another approach for diversifying a selected random peptide vector involves the mutagenesis of a pool, or subset, of recovered vectors. Recombinant host cells transformed with vectors recovered from panning are pooled and isolated. The vector DNA is mutagenized by treating the cells with, e.g.,

20   nitrous acid, formic acid, hydrazine, or by use of a mutator strain as described below. These treatments produce a variety of mutations in the vector DNA. The segment containing the sequence encoding the variable peptide can optionally be isolated by cutting with restriction nuclease(s) specific for sites flanking

25   the variable region and then recloned into undamaged vector DNA. Alternatively, the mutagenized vectors can be used without recloning of the mutagenized random peptide coding sequence.

In the second general approach for diversifying a set of peptide ligands, that of adding additional amino acids to a

30   peptide or peptides found to be active, a variety of methods are available. In one, the sequences of peptides selected in early panning are determined individually and new oligonucleotides, incorporating all or part of the determined sequence and an adjoining degenerate sequence, are synthesized. These are then

35   cloned to produce a secondary library.

In another approach that adds a second variable region to a pool of random peptide expression vectors, a restriction site is installed next to the primary variable region.

Preferably, the enzyme should cut outside of its recognition sequence, such as _BspMI_, which cuts leaving a four base 5' overhang, four bases to the 3' side of the recognition site. Thus, the recognition site may be placed four bases from the primary degenerate region. To insert a second variable region, a degenerately synthesized oligonucleotide is then ligated into this site to produce a second variable region juxtaposed to the primary variable region. This secondary library is then amplified and screened as before.

While in some instances it may be appropriate to synthesize peptides having contiguous variable regions to bind certain receptors, in other cases it may be desirable to provide peptides having two or more regions of diversity separated by spacer residues. For example, the variable regions may be separated by spacers that allow the diversity domains of the peptides to be presented to the receptor in different ways. The distance between variable regions may be as little as one residue or as many as five to ten to up to about 100 residues. For probing a large binding site, one may construct variable regions separated by a spacer containing 20 to 30 amino acids. The number of spacer residues, when present, will preferably be at least two to three or more but usually will be less than eight to ten. An oligonucleotide library having variable domains separated by spacers can be represented by the formula: $(NNK)_y$-$(abc)_n$-$(NNK)_z$, where N and K are as defined previously (note that S as defined previously may be substituted for K); y + z is equal to about 5, 6, 7, 8, or more; a, b and c represent the same or different nucleotides comprising a codon encoding spacer amino acids; and n is up to about 20 to 30 codons or more.

The spacer residues may be somewhat flexible, comprising polyglycine, for example, to provide the diversity domains of the library with the ability to interact with sites in a large binding site relatively unconstrained by attachment to the DNA binding protein. Rigid spacers, such as, e.g., polyproline, may also be inserted separately or in combination with other spacers, including glycine residues. The variable domains can be close to one another with a spacer serving to orient the one variable domain with respect to the other, such as

by employing a turn between the two sequences, as might be provided by a spacer of the sequence Gly-Pro-Gly, for example. To add stability to such a turn, it may be desirable or necessary to add Cys residues at either or both ends of each variable
5   region. The Cys residues would then form disulfide bridges to hold the variable regions together in a loop, and in this fashion may also serve to mimic a cyclic peptide. Of course, those skilled in the art will appreciate that various other types of covalent linkages for cyclization may also be accomplished.

10   The spacer residues described above can also be encoded on either or both ends of the variable nucleotide region. For instance, a cyclic peptide coding sequence can be made without an intervening spacer by having a Cys codon on both ends of the random peptide coding sequence. As above, flexible spacers,
15   e.g., polyglycine, may facilitate interaction of the random peptide with the selected receptors. Alternatively, rigid spacers may allow the peptide to be presented as if on the end of a rigid arm, where the number of residues, e.g., Pro, determines not only the length of the arm but also the direction for the arm
20   in which the peptide is oriented. Hydrophilic spacers, made up of charged and/or uncharged hydrophilic amino acids, (e.g., Thr, His, Asn, Gln, Arg, Glu, Asp, Met, Lys, etc.), or hydrophobic spacers made up of hydrophobic amino acids (e.g., Phe, Leu, Ile, Gly, Val, Ala, etc.) may be used to present the peptides to
25   binding sites with a variety of local environments.

The present invention can be used to construct improved spacer molecules. For example, one can construct a random peptide library that encodes a DNA binding protein, such as the lac repressor or a cysteine depleted lac repressor (described
30   below), a random peptide of formula NNK_5 (sequences up to and including NNK_{10} or NNK_{15} could also be used), and a peptide ligand of known specificity. One would then screen the library for improved binding of the peptide ligand to the receptor specific for the ligand using the method of the present invention; fusion
35   proteins that exhibit improved specificity would be isolated together with the vector that encodes them, and the vector would be sequenced to determine the structure of the spacer responsible for the improved binding.

Unless modified during or after synthesis by the translation machinery, recombinant peptide libraries consist of sequences of the 20 normal L-amino acids. While the available structural diversity for such a library is large, additional
5   diversity can be introduced by a variety of means, such as chemical modifications of the amino acids. For example, as one source of added diversity a peptide library of the invention can be subjected to carboxy terminal amidation. Carboxy terminal amidation is necessary to the activity of many naturally
10. occurring bioactive peptides. This modification occurs _in vivo_ through cleavage of the N-C bond of a carboxy terminal Gly residue in a two-step reaction catalyzed by the enzymes peptidylglycine alpha-amidation monooxygenase (PAM) and hydroxyglycine aminotransferase (HGAT). See, Eipper _et al._,
15   1991, _J. Biol. Chem._ _266_:7827-7833; Mizuno _et al._, 1986, _Biochem._ _Biophys._ _Res._ _Comm._ _137_(3): 984-991; Murthy _et al._, 1986, _J._ _Biol._ _Chem._ _261_(4): 1815-1822; Katopodis _et al._, 1990, _Biochemistry_ _29_:6115-6120; and Young and Tamburini, 1989, _J. Am._ _Chem._ _Soc._ _111_:1933-1934, each of which are incorporated herein
20   by reference.

Amidation can be performed by treatment with enzymes, such as PAM and HGAT, _in vivo_ or _in vitro_, and under conditions conducive to maintaining the structural integrity of the fusion protein/vector complex. In a random peptide library of the
25   present invention, amidation will occur on a library subset, i.e., those peptides having a carboxy terminal Gly. A library of peptides designed for amidation can be constructed by introducing a Gly codon at the end of the variable region domain of the library. After amidation, an enriched library serves as a
30   particularly efficient source of ligands for receptors that preferentially bind amidated peptides. Many of the C-terminus amidated bioactive peptides are processed from larger pro-hormones, where the amidated peptide is flanked at its C-terminus by the sequence -Gly-Lys-Arg-X . . . (SEQ. ID NO. 67)
35   (where X is any amino acid). Oligonucleotides encoding the sequence -Gly-Lys-Arg-X-Stop (SEQ. ID NO. 67) can be placed at the 3' end of the variable oligonucleotide region. When expressed, the Gly-Lys-Arg-X (SEQ. ID NO. 67) is removed by _in_

vivo or in vitro. enzymatic treatment, and the peptide library is carboxy terminal amidated as described above.

Conditions for C-terminal amidation of the libraries of the invention were developed using a model system that employed an antibody specific for the amidated C-terminus of the peptide cholecystokinin (CCK). The reaction conditions to make the peptide alpha-amidating monooxygenase (PAM) enzyme active when used to amidate the libraries were developed using an ^{125}I labeled small peptide substrate and an ELISA with a positive control glycine extended CCK octamer peptide fused to the lac repressor. The E. coli strain used in the experiment carried plasmid pJS129, which encodes the cysteine free lac repressor (described below) fused to the CCK substrate peptide (DYMGWMDFG) (SEQ. ID NO. 79).

A panning lysate was made from this strain using the standard panning protocol (see Example 6). After concentration of the column fractions in a Centriprep 100, four samples were prepared, each containing 0.25 ml of lysate and 0.25 ml of 2x PAM buffer (prepared by mixing 0.2 ml of 1 M HEPES, pH 7.4 (with KOH), 0.9 ml of 20% lactose, 3.65 ml of H_2O, and 0.1 ml of a solution composed of 20 mg/ml catalase, 100 μl of 20 mM $CuSO_4$, 16.6 μl of 6 M NaI, and 150 μl of 0.1 M ascorbic acid). PAM enzyme was added to the tubes in different amounts and incubated at 37°C for 30 minutes. Then, 120 μl of 5% BSA in HEKL buffer and 6 μl of 10 mg/ml herring DNA were added to each tube; the contents of each tube were then added to 6 microtiter wells that had been coated with 2 μg/well anti-CCK antibody and blocked with BSA. The microtiter plate was agitated at 4°C for 150 minutes, washed 5x with cold HEKL, washed for 10 minutes with a solution composed of HEKL, 1% BSA, and 0.1 mg/ml herring DNA, and washed again 5x with cold HEKL. The plasmids were eluted using the standard protocol and used to transform E. coli host cells. The results showed a dramatic increase in the recovery of plasmid transformants with increasing amounts of PAM enzyme, demonstrating that the amidation reaction worked.

Other modifications found in naturally occurring peptides and proteins can be introduced into the libraries to provide additional diversity and to contribute to a desired biological activity. For example, the variable region library

can be provided with codons that code for amino acid residues involved in phosphorylation, glycosylation, sulfation, isoprenylation (or the addition of other lipids), etc. Modifications not catalyzed by naturally occurring enzymes can be

5   introduced by chemical means (under relatively mild conditions) or through the action of, e.g., catalytic antibodies and the like.   In most cases, an efficient strategy for library construction involves specifying the enzyme (or chemical) substrate recognition site within or adjacent to the variable

10   nucleotide region of the library so that most members of the library are modified.   The substrate recognition site added can be simply a single residue (e.g., serine for phosphorylation) or a complex consensus sequence, as desired.

Conformational constraints, or scaffolding, can also be

15   introduced into the structure of the peptide libraries. A number of motifs from known protein and peptide structures can be adapted for this purpose.   The method involves introducing nucleotide sequences that code for conserved structural residues into or adjacent to the variable nucleotide region so as to

20   contribute to the desired peptide structure.   Positions nonessential to the structure are allowed to vary.

A degenerate peptide library as described herein can incorporate the conserved frameworks to produce and/or identify members of families of bioactive peptides or their binding

25   receptor elements.   Several families of bioactive peptides are related by a secondary structure that results in a conserved "framework," which in some cases is a pair of cysteines that flank a string of variable residues.   This results in the display of the variable residues in a loop closed by a disulfide bond, as

30   discussed above.

In some cases, a more complex framework that contributes to the bioactivity of the peptides is shared among members of a peptide family.   An example of this class is the conotoxins: peptide toxins of 10 to 30 amino acids produced by

35   venomous molluscs known as predatory cone snails.   The conotoxin peptides generally possess a high density of disulfide crosslinking.   Of those that are highly crosslinked, most belong

to two groups, mu and omega, that have conserved primary
frameworks as follows (C is Cys):

     mu   CC . . . . . C . . . . . C . . . . . CC; and
5   omega   C . . . . . C . . . . . CC . . . . . C C

The number of residues flanked by each pair of Cys residues
varies from 2 to 6 in the peptides reported to date. The side
chains of the residues that flank the Cys residues are apparently
10 not conserved in peptides with different specificity, as in
peptides from different species with similar or identical
specificities. Thus, the conotoxins have exploited a conserved,
densely crosslinked motif as a framework for hypervariable
regions to produce a huge array of peptides with many different
15 pharmacological effects.

The mu and omega classes (with 6 Cys residues) have 15
possible combinations of disulfide bonds. Usually only one of
these conformations is the active ("correct") form. The correct
folding of the peptides may be directed by a conserved 40 residue
20 peptide that is cleaved from the N-terminus of the conopeptide to
produce the small, mature bioactive peptides that appear in the
venom.

With 2 to 6 variable residues between each pair of Cys
residues, there are 125 (5^3) possible framework arrangements for
25 the mu class (2,2,2, to 6,6,6), and 625 (5^4) possible for the
omega class (2,2,2,2 to 6,6,6,6). Randomizing the identity of
the residues within each framework produces 10^{10} to >10^{30}
peptides. "Cono-like" peptide libraries are constructed having
a conserved disulfide framework, varied numbers of residues in
30 each hypervariable region, and varied identity of those residues.
Thus, a sequence for the structural framework for use in the
present invention comprises Cys-Cys-Y-Cys-Y-Cys-Cys, or
Cys-Y-Cys-Y-Cys-Cys-Y-Cys-Y-Cys, where Y is $(NNK)_x$ or $(NNS)_x$; N
is A, C, G or T; K is G or T; S is G or C; and x is from 2 to 6.
35 Framework structures that require the formation of one
or more disulfide bonds under oxidizing conditions may create
problems with respect to the natural _lac_ repressor, which has 3
cysteine residues. All 3 of these residues, however, can be

changed to other amino acids without a serious effect on the function of the molecule (see Kleina and Miller, supra). Plasmid pJS123 is derived from plasmid pMC5 by site specific mutagenesis and encodes a lac repressor identical to the lac repressor

5   encoded on plasmid pMC5, except the cysteine codon at position 107 has been changed to an serine codon; the cysteine codon at position 140 has been changed to an alanine codon (alanine works better than serine at this position); and the cysteine codon at position 281 has been changed to a serine codon. Plasmid pJS123

10   (available in strain ARI 161 from the American Type Culture Collection under the accession number ATCC No. 68819) is therefore preferred for constructing random peptide libraries involving cysteine-linked framework structures.

The lac repressor coding sequence in plasmid pJS123 can

15   be subjected to mutagenesis to improve the binding of the mutant protein with lacO type sequences. A preferred method for performing this mutagenesis involves the construction of a coding sequence in plasmid pJS123 that encodes a fusion protein comprised of the cysteine depleted lac repressor, a spacer

20   peptide, and a peptide ligand of known specificity. The resulting vector is subjected to mutagenesis by any of a variety of methods; a preferred method involves transformation of an E. coli mutator strain such as mutD5 (see Schaaper, 1988, Proc. Natl. Acad. Sci. USA 85:8126-8130, incorporated herein by

25   reference) and culture of the transformants to produce the fusion protein encoded by the vector. The fusion proteins are screened by the present method to find vectors that have been mutated to increase the binding affinity of the cysteine depleted lac repressor to the lacO sequence. One could combine this method

30   with the method of constructing improved spacers, described above, to select for an improved cysteine depleted lac repressor-peptide spacer molecule.

In such a fashion, plasmid pJS123 was modified to include a D32.39 antibody epitope to create plasmid pJS128, which

35   was then introduced into a mutD mutator strain. Oligonucleotides were then cloned into the mutagenized vector to encode a D32.39 epitope joined to repressor via a random region of 5, 10, or 15 amino acids. This library was panned on D32.39 antibody for 5

rounds under increasingly stringent conditions. Individual clones were selected from the population of plasmids surviving after the fifth round and tested by a variety of assays. These assays included: (1) tests for ability to repress the
5 chromosomal lac operon (a test of DNA binding affinity); (2) tests for plasmid copy number; (3) ELISA with D32.39 antibody to test for display of the peptide epitope; and (4) tests of plasmid recovery during panning. Several of these plasmids were sequenced in the random tail region to determine the structure of
10 the linker peptide. A series of subcloning experiments were also conducted to determine regions of the plasmids that determined the observable properties of the plasmids. Finally, plasmids carrying a higher copy number replication origin and encoding one of the linker regions were constructed and sequenced to ascertain
15 that no base changes in the cysteine free repressor gene, as compared to the starting plasmid, were introduced. The linker tail of this plasmid and the cloning strategy for random libraries is shown in Figure 4. Two versions of the vector were constructed, one with the cysteine-free lac repressor gene
20 (ARI246/pJS141; ATCC No. 69088) and one with the wild-type lac repressor gene (ARI280/pJS142; ATCC No. 69087).

ARI246 has the genotype E. coli B lon-11 sulA1 hsdR17 Δ(ompT-fepC) ΔclpA319::kan lacI42::Tn10 lacZU118. The lon-11, Δ(ompT-fepC), and ΔclpA319::kan mutations destroy three genes
25 involved in proteolysis, so this strain should allow greater diversity of peptides to be expressed on the library particles. The sulA1 mutation suppresses the filamentation phenotype caused by the lon-11 allele. The hsdR17 mutation destroys the restriction system to allow more efficient transformation of
30 unmodified DNA. The lacI42::Tn10 mutation eliminates expression of the chromosomal lac repressor gene to prevent competition of wild-type repressor for binding sites on the library plasmids. The lacZU118 allele stops expression of β-galactosidase, which would otherwise be constitutive in the lacI42::Tn10 background,
35 leading to unnecessary use of cell resources and reducing growth rates. E. coli B cells grow more quickly than K12 cells and yield excellent electrocompetent cells for transformation. Transformation frequencies of around 5 x 10^{10} tf/μg of Bluescript

plasmid DNA can be frequently observed with ARI246 cells. ARI280 has the same genotype as ARI246, except that the lacI mutation has been converted to a deletion by selecting for loss of the Tn10 insertion, and a recA::cat mutation has been introduced. The recA::cat mutation is useful to prevent homologous recombination between plasmids. As a consequence, the library plasmids exist more frequently as monomers, rather than multimeric forms that can be observed in ARI246. The monomers are better for two reasons: monomers reduce the valency of peptides per library particle, allowing more stringent selection for higher affinity peptide ligands; and growth as monomers increases the number of plasmids per amount of DNA, increasing the number of library equivalents that can be panned against receptors. The recA::cat mutation makes the strain less healthy, so growth rates are slower, and the transformation frequency is reduced to about 2×10^{10} tf/μg.

Other changes can be introduced to provide residues that contribute to the peptide structure, around which the variable amino acids are encoded by the library members. For example, these residues can provide for alpha helices, a helix-turn-helix structure, four helix bundles, a beta-sheet, or other secondary or tertiary structural (framework or scaffolding) motifs. See PCT patent publication no. 91/19818, incorporated herein by reference. DNA binding peptides, such as those that correspond to the transcriptional transactivators referred to as leucine zippers, can also be used as a framework, provided the DNA binding peptide is distinct from the DNA binding protein component of the fusion protein and the library vector does not contain the binding site for the DNA binding peptide. In these peptides, leucine residues are repeated every seven residues in the motifs, and the region is adjacent to an alpha helical region rich in lysines and arginines and characterized by a conserved helical face and a variable helical face.

Other specialized forms of structural constraints can also be used in the present invention. For example, certain serine proteases are inhibited by small proteins of conserved structure (e.g., pancreatic trypsin inhibitor). This conserved framework can incorporate degenerate regions as described herein

to generate libraries for screening for novel protease inhibitors.

    In another aspect related to frameworks for a peptide library, information from the structure of known ligands can be
5   used to find new peptide ligands having features modified from those of the known ligand. In this embodiment, fragments of a gene encoding a known ligand, prepared by, e.g., limited DNAse digestion into pieces of 20 to 100 base pairs, are subcloned into a variable nucleotide region system as described herein either
10 singly or in random combinations of several fragments. The fragment library is then screened in accordance with the procedures herein for binding to the receptor to identify small peptides capable of binding to the receptor and having characteristics which differ as desired from the parental peptide
15 ligand. This method is useful for screening for any receptor-ligand interaction where one or both members are encoded by a gene, e.g., growth factors, hormones, cytokines and the like, such as insulin, interleukins, insulin-like growth factor, etc. In this embodiment of the invention, the peptide library
20 can contain as few as 10 to 100 different members, although libraries of 1,000 or more members will generally be used.

    Thus, the present invention can be used to construct peptide ligands of great diversity. The novel features of the preferred embodiment of the invention, called "peptides on
25 plasmids", in which the lac repressor is the DNA binding protein and a plasmid vector encodes the fusion protein, are distinct from those of the previously described phage libraries. The random peptides of the present libraries can be displayed with a free carboxy terminus instead of being displayed at the amino
30 terminus or internal to the carrier protein and so add diversity to the peptide structures available for receptor binding. The presentation of peptide ligands at the carboxy terminus also facilitates amidation, as discussed above. This mode of display also ensures that stop codons in the degenerate region, which
35 occur more often in longer degenerate oligonucleotides, shorten rather than destroy individual clones. The presence of stop codons in the random peptide coding sequence actually serves to create additional diversity, by creating peptides of differing

lengths. The _lac_ repressor fusions of the invention also allow the display of potential ligands with a wide range of sizes.

In addition, these _lac_ repressor fusions are cytoplasmic proteins, unlike the phage fusions, which are exported to the periplasm. The use of both fusion methods increases total available peptide diversity, because the two types of libraries are exposed to different cellular compartments and so are exposed to different sets of _E. coli_ proteases and to different reduction/oxidation environments. There is no need, however, for peptides fused to the _lac_ repressor to be compatible with the protein export apparatus and the formation of an intact phage coat. The peptides need simply be compatible with the formation of at least a repressor dimer, which is the smallest form of the protein that can bind DNA (see Daly and Matthews, 1986, _Biochem._ 25:5474-5478, and Kania and Brown, 1976, _Proc. Natl. Acad. Sci. USA_ 73:3529-3533).

As in the phage system, the _lac_ repressor fusion library displays multiple copies of the peptide on each library particle. Each repressor tetramer, in principle, displays four peptides that are available for binding to receptors. In addition, each plasmid monomer can bind up to two tetramers (if no loop is formed), and multimers of the plasmid can display higher multiples of two tetramers. This multivalent display allows the isolation of ligands with moderate affinity (micromolar K_d, see Cwirla _et al._, _supra_). For receptors with known, high affinity peptide ligands, these moderate affinity ligands can obscure the high affinity ones simply because of their greater numbers. This problem can be overcome by immobilizing monovalent receptors at low density, which allows high affinity (nanomolar K_d) ligands to be identified, as discussed above. For receptors whose normal ligands are not small peptides, however, this multivalency of display will be an advantage for identifying initial families of moderate affinity ligands, which can then be optimized by additional rounds of screening under monovalent conditions. The multivalency of ligand display therefore allows the isolation of peptides with a wide range of affinities, depending on the density of the receptor during the panning procedure.

Libraries of peptides produced and screened according to the present invention are particularly useful for mapping antibody epitopes. The ability to sample a large number of potential epitopes as described herein has clear advantages over the methods based on chemical synthesis now in use and described in, among others, Geysen et al., 1987, J. Immunol. Meth. 102:259-274. In addition, these libraries are useful in providing new ligands for important binding molecules, such as hormone receptors, adhesion molecules, enzymes, and the like.

The present libraries can be generalized to allow the screening of a wide variety of peptide and protein ligands. In addition, the vectors are constructed so that screening of other ligands encoded by the plasmid is possible. For example, the system can be simply modified to allow screening of RNA ligands. A known RNA binding protein (e.g., a ribosomal protein) is fused to the DNA binding protein. A promoter elsewhere on the vector drives expression of an RNA molecule composed of the known binding site for the RNA binding protein followed by random sequence. The DNA-RNA binding fusion protein would link the genetic information of the vector with each member of a library of RNA ligands. These RNA ligands could then be screened by panning techniques.

Another large class of possible extensions to this technique is to use a modified version of the vector to isolate genes whose products modify peptides, proteins, or RNA in a desired fashion. This requires the availability of a receptor that binds specifically to the modified product. For the general case, a connection is made between the plasmid and the substrate peptide, protein, or RNA, as described above. The plasmid is then used as a cloning vector to make libraries of DNA or cDNA from a source with the potential to contain the desired modification gene (specific organisms, PCR amplified antibody genes, etc.) under the control of a promoter that functions in E. coli. Plasmids carrying the gene in question could then be isolated by panning lysates of the library with the receptor specific for the modified product.

For example, a gene encoding an enzyme that cleaves a particular amino acid sequence could be isolated from libraries

of DNA from organisms that might have such a protease or from amplified antibody cDNA. An antibody for use as the receptor would first be made to the peptide that would remain after the desired cleavage reaction had taken place. Many such antibodies

5   will not bind to that peptide unless it is exposed at the N- or C-termius of the protein. The coding sequence for the uncleaved substrate sequence would be attached to the DNA binding protein coding sequence in a vector. This vector would be used to make an expression library from an appropriate source. Members of

10   this library containing a gene that encoded an enzyme able to cleave the peptide would cleave only the peptide attached to the plasmid with that gene. Panning of lysates of the library would preferentially isolate those plasmids with active genes.

     As can be appreciated from the disclosure above, the present

15   invention has a wide variety of applications. Accordingly, the following examples are offered by way of illustration, not by way of limitation.

EXAMPLE 1

Construction of Plasmids pMC3 and pMC5

20   The bacterial strains used were E. coli K12 strains MC1061 (araD139 Δ(araABC-leu)7696 thr ΔlacX74 galU galK hsdR mcrB rpsL(strA) thi), ARI 20 (F' lac⁺ pro⁺ lacIqL8 lacIam74 // Δ(lac-pro) thi rpsL(strA) recA::cat), and XL1-Blue (F' proAB lacIq lacZΔM15 Tn10 // recA1 endA1

25   gyrA96 thi hsdR17 supE44 relA1 lac), and E. coli B strain ARI 161 (lon-11, sulA1, hsdR17, Δ(ompT-fepC), ΔclpA319::kan). ARI 161 is a protease deficient strain and serves to minimize proteolysis of the peptides in the library, which would reduce the available diversity for panning. Mutations known to reduce

30   proteolysis include degP, lon, htpR, ompT, and clpA,P.

         The library plasmid pMC5 was constructed in several steps using plasmid pBAD18 as the starting plasmid. Plasmid pBAD18 contains the araB promoter followed by a polylinker and a terminator under the control of the positive/negative regulator

35   AraC, also specified by the plasmid. Plasmid pBAD18 also contains a modified plasmid pBR322 origin and the bla gene to permit replication and selection in E. coli, as well as the phage

M13 intragenic region to permit rescue of single-stranded DNA for sequencing.

The lacI gene was modified for cloning into plasmid pBAD18 using the GeneAmp® PCR amplification kit (Perkin-Elmer
5 Cetus Instruments) with oligonucleotides ON-286 and ON-287, shown below:

ON-286   5'-GCG GGC TAG CTA ACT AAT GGA GGA TAC ATA AAT GAA
            ACC AGT AAC GTT ATA CG-3' (SEQ. ID NO. 68)
10 ON-287   5'-CGT TCC GAG CTC ACT GCC CGC TCT CGA GTC GGG AAA CCT
            GTC GTG C-3' (SEQ. ID NO. 69).

The amplification reaction was carried out according to the manufacturer's instructions, except for the use of Vent™ DNA
15 polymerase (New England Biolabs). ON-286 contains a nonhomologous 5' region that adds an NheI site, a consensus ribosome binding site (see Gold and Stormo, 1990, Methods in Enzymology (Goeddel, ed., Boston: Academic Press), pp. 89-103, incorporated herein by reference), and changes the initiation
20 codon of lacI from GTG to ATG. ON-287 changes codons 356 and 357 of lacI to an XhoI site through two silent mutations, and adds a SacI site after the lacI stop codon.

Cloning of the NheI, SacI digested amplification product into plasmid pBAD18 produced vector pJS100. Two lacO$_S$
25 sequences were added to this vector, with their centers spaced 326 bp apart, by amplifying an unrelated sequence (the human D$_2$ dopamine receptor gene (see England et al., 1991, FEBS Lett. 279:87-90, incorporated herein by reference), with oligonucleotides ON-295 and ON-296, shown below:
30 ON-295   5'-CCT CCA TAT GAA TTG TGA GCG CTC ACA ATT CGG TAC AGC
            CCC ATC CCA CCC-3' (SEQ. ID NO. 70)
    ON-296   5'-CGC CAT CGA TCA ATT GTG AGC GCT CAC AAT TCA GGA TGT
            GTG TGA TGA AGA-3' (SEQ. ID NO. 71)
    ON-295 adds an NdeI site and a lacO$_S$ sequence at one end of the
35 amplified fragment, and ON-296 adds a ClaI site and lacO$_S$ at the other end. Cloning of the NdeI to ClaI fragment into pJS100 produced plasmid pJS102.

Plasmid pMC3, encoding the dynorphin B-tailed <u>lac</u> repressor, was constructed by cloning complementary oligonucleotides ON-312 and ON-313 to replace the <u>Xho</u>I to <u>Xba</u>I fragment at the 3' end of <u>lacI</u> in pJS102. These oligonucleotides

5   add sequence encoding a five amino acid spacer (GADGA [SEQ. ID NO.:65]) and dynorphin B (YGGFLRRQFKVVT [SEQ. ID NO.:66]) to the end of the wild-type <u>lacI</u> sequence, introduce an <u>Sfi</u>I site in the sequence encoding the spacer, and are shown below:

ON-312   5'-TCG AGA GCG GGC AGG GGG CCG ACG GGG CCT ACG GTG GTT

10   TCC TGC GTC GTC AGT TCA AAG TTG TAA CCT AAT-3'

(SEQ. ID NO. 72)

ON-313   5'-CTA GAT TAG GTT ACA ACT TTG AAC TGA CGA CGC AGG

               AAA CCA CCG TAG GCC CCG TCG GCC CCC TGC CGG CTC-3'

(SEQ. ID NO. 73)

15   The library plasmid pMC5 was constructed by cloning complementary oligonucleotides ON-335 and ON-336 to replace the <u>Sfi</u>I to <u>Hin</u>dIII dynorphin B segment of pMC3, as shown in Figure 2. Oligonucleotides ON-335 and ON-336 are shown below:

20   ON-335   5'-GGG CCT AAT TAA TTA-3' (SEQ. ID NO. 74)

ON-336   5'-AGC TTA ATT AAT TAG GCC CCG T-3' (SEQ. ID NO. 75)

Plasmid pMC3 is available in strain ARI161 from the American Type Culture Collection under the accession number ATCC No. 68818.

25

EXAMPLE 2

<u>Construction of a Random Dodecamer Peptide Library</u>

Oligonucleotide ON-332 was synthesized with the sequence:

30   5'-GT GGC GCC (NNK)$_{12}$ TAA GGT CTC G-3', (SEQ. ID NO. 76) where N is A,C,G, or T (equimolar) and K is G or T (see Cwirla <u>et al</u>., supra). The oligonucleotide was purified by HPLC and phosphorylated with T4 kinase (New England Biolabs). The two half-site oligonucleotides ON-369 and ON-370 were phosphorylated

35   during synthesis and are shown below:

ON-369   5'-GGC GCC ACC GT-3' (SEQ. ID NO. 77)

ON-370   5'-AGC TCG AGA CCT TA-3' (SEQ. ID NO. 78)

ON-369 and ON-370 annealed to ON-332 produce SfiI and HindIII-compatible ends, respectively, but the ligated product does not have either recognition sequence (see Figure 2).

Four hundred pmoles of each oligonucleotide were
5   annealed in a 25 µl reaction buffer (10 mM Tris, pH 7.4, 1 mM EDTA, 100 mM NaCl), by heating to 65°C for 10 min. and cooling for 30 min. to room temperature. Vector pMC5 was digested to completion with SfiI and HindIII, the vector backbone was isolated by 4 rounds of washing with TE buffer (10 mM Tris, pH
10   8.0, 1 mM EDTA) in a Centricon 100 microconcentrator (Amicon) by the manufacturer's instructions, followed by phenol extraction and ethanol precipitation. The annealed oligonucleotides were added to 64 micrograms of digested pMC5 at a 4:1 molar ratio in a 3.2 ml ligation reaction containing 5% PEG, 3200 units of
15   HindIII, 194 Weiss units of T4 ligase (New England Biolabs), 1mM ATP, 20 mM Tris, pH 7.5, 10 mM MgCl2, 0.1mM EDTA, 50 µg/ml BSA, and 2 mM DTT. The reaction was split equally into 8 tubes and incubated overnight at 15°C.

After ethanol precipitation, 1/16 of the ligated DNA (4
20   µg) was introduced into MC1061 (80 µl) by electroporation (Dower et al., 1988, Nucl. Acids Res. 16:6127-6145, incorporated herein by reference), to yield 5.5 x 10^8 independent transformants. The library was amplified approximately 1000-fold in 1 liter of LB/100 µg/ml ampicillin by growth of the transformants at 37°C to
25   an A$_{600}$ of 1. The cells containing the library were concentrated by centrifugation at 5500 x g for 6 min., washed once in ice-cold 50 mM Tris (pH 7.6), 10 mM EDTA, 100 mM KCl, followed by a wash in ice-cold 10 mM Tris, 0.1 mM EDTA, 100 mM KCl. The final pellet was resuspended in 16 ml of HEG buffer (35 mM HEPES/KOH pH
30   7.5, 0.1 mM EDTA, 100 mM Na Glutamate), distributed into 19 tubes of 1.0 ml each, frozen on dry ice, and stored at -70°C.

EXAMPLE 3

Panning the Library

35   One aliquot (1.0 ml) of the library prepared in Example 2 was thawed on ice and added to 9 ml of lysis buffer (35 mM HEPES {pH 7.5 with KOH}, 0.1 mM EDTA, 100 mM Na glutamate, 5% glycerol, 0.3 mg/ml BSA, 1 mM DTT, and 0.1 mM PMSF). Lysozyme

was added (0.3 ml at 10 mg/ml in HEG), and the mixture was incubated on ice for 1 hr.

The cellular debris was removed by centrifugation of the lysate at 20,000 x g for 15 min., and the supernatant was concentrated by centrifugation in a Centriprep® 100 concentrator (Amicon) at 500 x g for 40 min. The concentrated supernatant (about 0.5 ml) was washed with 10 ml of HEG buffer and centrifuged as before. A sample (5%) of the total lysate was removed to determine the pre-panned input of plasmid complexes.

An alternate method for partially purifying and concentrating the lysate is as follows. About 2.0 ml of the frozen cells in HEG are thawed on ice, and then 8 ml of lysis buffer without Na glutamate (high ionic strength inhibits lysozyme; DTT is optional) are added to the cells, and the mixture is incubated on ice for 1 hr. The cellular debris is removed from the lysate by centrifugation at 20,000 x g for 15 min., and the supernatant is loaded onto a Sephacryl® S-400 High Resolution (Pharmacia) gel-filtration column (22 mm x 250 mm). The plasmid-fusion protein complexes elute in the void volume. The void volume (30 ml) is concentrated with two Centriprep® 100 concentrators, as described above. After adjusting the Na glutamate concentration of the concentrate, one carries out the remainder of the procedure in the same manner as with the first method.

Half of the remaining concentrated lysate was added to D32.39-antibody-coated sheep-anti-mouse (Fc)-coupled magnetic beads (10 μg of D32.39 added to 5 mg Dynal beads for 1 hr. at 25°C followed by 6 washes with HEG), and half was added to uncoated beads. After incubating the lysates with the beads at 0°C for 1 hr. with shaking, the beads were washed three times with 5 ml of cold HEG/0.1% BSA and then three times with HEG using a MACS 0.6 tesla magnet (Miltenyi Biotec GmBH) to immobilize the beads. The plasmids were dissociated from the beads by phenol extraction, and after adding 20 μg of glycogen (Boehringer Mannheim), the DNA was precipitated with an equal volume of isopropanol. The pellet was washed with 75% ethanol, and the DNA was resuspended in either 4 μl (panned DNA) or 400 μl (pre-panned DNA) of H_2O. Strain MC1061 was transformed using 2 μl

each of the DNA solutions to permit counts of recovered plasmids and amplification of the selected plasmids. The results of the panning are shown below in Table 1.

Table 1

| Panning Round | Input | Number of Transformants | |
		Ab D32.39 Beads	Uncoated Beads
1	1.6×10^{10}	9×10^7	1.7×10^5
2	1.4×10^{11}	6.1×10^7	1.2×10^4
3	1.7×10^{11}	2.0×10^9	40
4	----	1.6×10^{11}	4×10^4

EXAMPLE 4

ELISA Analysis of the Library

An ELISA was used to test MC1061 transformants from the second, third, and fourth rounds for D32.39-specific ligands (see Example 3). The ELISA was performed in a 96-well plate (Beckman). Single colonies of transformants obtained from panning were grown overnight in LB/100 μg/ml ampicillin at 37°C. The overnight cultures were diluted 1/10 in 3 ml LB/100 μg/ml ampicillin and grown 1 hr. The expression of the _lac_ repressor-peptide fusions was induced by the addition of arabinose to a final concentration of 0.2%.

The cells were lysed as described above in 1 ml of lysis buffer plus lysozyme and stored at -70°C. Thawed crude lysate was added to each of 2 wells (100 μl/well), and the plate was incubated at 37°C. After 45 min, 100 μl of 1% BSA in PBS (10 mM NaPO4, pH 7.4, 120 mM NaCl, and 2.7 mM KCl) were added for an additional 15 min. at 37°C, followed by 3 washes with PBS/0.05% Tween 20. Each well then was blocked with 1% BSA in PBS (200 μl/well) for 30 min. at 37°C, and the wells were washed as before.

The primary antibody, D32.39 (100 μl of antibody at 1 μg/ml in PBS/0.1% BSA) was added to each well, the plate was incubated at room temperature for 1 hr., and then each well was

washed as before. The secondary antibody, alkaline phosphatase-conjugated Goat-anti-rabbit antibody (Gibco-BRL), was diluted 1/3000 in PBS/0.1% BSA and added to each well (100 µl/well); the plate was then incubated for 1 hr at room
5   temperature. After three washes with PBS/0.05% Tween 20 and two with TBS (10 mM Tris pH7.5, 150 mM NaCl), the ELISA was developed with 4 mg/ml p-nitrophenyl phosphate in 1 M diethanolamine/HCl pH 9.8, 0.24 mM MgCl$_2$ (200 µl/well).

The reaction was stopped after 6 min. by the addition
10 of 2 M NaOH (50 µl/well), and the absorbance at 405 nm was measured on a plate reader (a Biomek, from Beckman). The positive control for the ELISA was MC1061 transformed with pMC3, encoding the <u>lac</u> repressor-dynorphin B fusion. The negative controls were wells not coated with lysate. Background
15 variability was calculated from the wells containing lysates from 16 colonies selected at random from the library, none of which scored significantly above the negative controls. Wells were scored as positive if the measured absorbance was at least two standard deviations above background.

20 Of randomly picked colonies, 35 of 58 (60%) tested positive by ELISA: 11 of 20 from round two, 12 of 16 from round three, and 12 of 22 from round four. None of 16 random colonies from the unpanned library scored significantly above background. These data demonstrate the rapid enrichment of specific ligands
25 achieved by the present invention: after only two rounds of panning, the majority of plasmids encoded peptides with affinity for the D32.39 antibody.

To determine the structure of the peptide ligands obtained by the present method, plasmids from both ELISA positive
30 and ELISA negative colonies obtained after panning were sequenced. Double stranded plasmid DNA, isolated from strain XL1-Blue, was sequenced using Sequenase® (US Biochemicals) according to the instructions supplied by the manufacturer.

The translated peptide sequence for all ELISA positive
35 colonies examined shared the consensus sequence shown in Figure 3. The preferred recognition sequence for the D32.39 antibody apparently covers a six amino acid region of the dynorphin B peptide (RQFKVV) (SEQ. ID NO. 80). In the first position,

arginine is invariant for all of the ELISA positive clones. No
strong bias was evident for residues in the second position. In
the third position, however, five amino acids (phenylalanine,
histidine, asparagine, tyrosine, and tryptophan, in order of
5   frequency) account for 98% of the residues. Of these, the
aromatic amino acids comprise 74% of this total. The fourth
position shows a strong bias for the positively charged residues
lysine (69%) and arginine (21%). The fifth position is occupied
almost exclusively by hydrophobic residues, most of which are
10   valine (81%). Valine and threonine predominate in the sixth
position (76%), with serine and isoleucine accounting for most of
the remaining amino acids.

Of the ELISA negative clones obtained after panning,
greater than half showed peptide sequence similarity to the
15   consensus motif (Figure 3). None of 19 isolates sequenced from
the unpanned library showed any such similarity. Some of these
ELISA negative sequences differ enough from the consensus that
their affinity for the antibody may be insufficient to permit
detection in the ELISA. There are, however, ELISA negative
20   sequences identical in the five conserved amino acids of the
consensus region to clones that scored positive (e.g., #28 and
#57). There may be amino acids outside the consensus region that
affect binding of the peptide to antibody or its susceptibility
to E. coli proteases, or its availability in the ELISA. That
25   even the ELISA negative clones frequently have an obvious
consensus sequence demonstrates the utility of the present
invention for isolating ligands for biological receptors.

Example 5

30   Construction of Plasmid pDimer1

Plasmid pDimer1 was constructed to provide a monovalent
display system. The plasmid encodes a lac headpiece dimer linked
to a dynorphin B antibody epitope (RQFKVVT) (SEQ. ID NO. 81); a
random linker is encoded between the headpieces and between the
35   second headpiece and the dynorphin B epitope. This vector was
used in the affinity enrichment protocol with anti-dynorphin B
antibody to determine the optimum linker sequences, as discussed
below.

Plasmid pDimer1 was constructed from plasmid pMC5 by first using plasmid pMC5 as a PCR template for amplification with oligonucleotide primers ON-929 and ON-930. These oligonucleotides are shown below.

5

ON-929:   5'-TATTTGCACGGCGTCACACTT-3' (SEQ. ID NO. 82)

ON-930:

5'-CCGCGCCTGGGCCCAGGGAATGTAATTGAGCTCCGCCATCGCCGCTT-3' (SEQ. ID NO. 83)

10

The amplified DNA resulting from the PCR amplification was digested with restriction enzymes BamHI and ApaI and then ligated with the large BamHI-ApaI fragment of plasmid pMC5 to yield plasmid pMC5dlacI. This process resulted in a deletion of all but about the first 60 codons (the lac headpiece coding sequence) of the lacI gene coding sequence from plasmid pMC5.

15

To generate the second headpiece-encoding region, plasmid pMC5 was used as a template for PCR amplification with oligonucleotide primers ON-938 and ON-940 (to encode a five amino acid random linker between the two headpieces) and ON-939 and ON-940 (to encode a four amino acid random linker between the two headpieces). These oligonucleotides are shown below (M is A or C).

20

25 ON-938:

5'-CGATGGCGGAGCTCAATTACATTCCCNNKNNKNNKNNKNNKAAACC-AGTAACGTTATACGAT-3' (SEQ. ID NO. 84)

ON-939:

30 5'-CGATGGCGGAGCTCAATTACATTCCCNNKNNKNNKNNKAAACCAGT-AACGTTATACGAT-3' (SEQ. ID NO. 85)

ON-940:

5'-CGCCCGCCAAGCTTAGGTTACAACTTTGAACTGACGMNNMNNMNNM-NNGGGAATGTAATTCAGCTCCGCCAT-3' (SEQ. ID NO. 86)

35

The amplified DNA from each reaction was digested with restriction enzymes SstI and HindIII and then ligated to the

large <u>Sst</u>I-<u>Hind</u>III fragment of plasmid pMC5dlacI. The resulting
plasmids were designated pDimer1 and were then transformed into
<u>E</u>. <u>coli</u> host cells as a mixture.

    The resulting library of transformants was panned in
5 accordance with the procedure of Example 6, except that HEK
·buffer rather than HEKL buffer was used in the lysis and column
purification steps. The cells were grown either in LB broth with
0.1% glucose and supplemented with L-arabinose to 0.2% about 30
minutes before harvesting the plasmids or in LB broth without
10 glucose and supplemented with L-arabinose to 0.2% about 15
minutes before harvesting the plasmids. Both protocols yielded
similar results. Four rounds of affinity enrichment were
completed; in the latter two rounds, about 5 to 10 μg/ml of an
oligonucleotide encoding <u>lacO</u>$_S$ was added during the lysis and
15 panning steps to increase the stringency of selection for tight-
binding derivatives. This oligonucleotide is shown below in
single-stranded form but is used in double-stranded form.

ON-413: 5'-GAATTCAATTGTGAGCGCTCACAATTGAATTC-3' (SEQ. ID NO. 87)
20

    After the fourth round of affinity enrichment, a number
of individual plasmids were sequenced to determine the identity
of the linker between the two headpieces and the second headpiece
and the dynorphin B sequence. Both four and five amino acid
25 linkers were observed between the two headpieces; about 25% of
the isolates examined had the sequence GRCR between the two
headpieces and the sequence GPNQ between the second headpiece and
the dynorphin B sequence. These results show that the two
headpiece system can be used to screen ligands for ability to
30 bind to receptors.

<div align="center">Example 6</div>

<u>Standard Protocol</u>
    This Example provides a standard protocol for the
35 method of the present invention with any receptor that can be
immobilized on a microtiter dish with an immobilizing antibody.
To practice the method, the following reagents will be helpful.

Items	Vendor	Catalog #
BSA, fraction V, RIA grade	USB	10868
BSA, protease free	USB	10867
Bulk DNA, sonicated, phenol extracted		
Centriprep 100 concentrator, 5-15 ml	Amicon	4308
Chromatography column, G22X250	Amicon	95220
Coomassie Plus protein assay reagent	Pierce	23236
DTT		
EDTA, disodium, dihydrate	Sigma	E-5134
Ethyl alcohol, 200 proof	Gold Shield Chem.	
Glycerol	Sigma	G-5516
Glycogen, molecular biology grade	Boehringer	901 393
HEPES free acid, molecular biology grade	Sigma	H-0891
Isopropanol, HPLC grade	Aldrich	27,049-0
IPTG	Bachem	SISO10
alpha-Lactose, monohydrate	Sigma	L-3625
Lysozyme, from hen egg white	Boehringer	837 059
Microtiter plate, Immulon 4, flat bottom	Dynatech	011-010-3850
PBS	Sigma	1000-3
PMSF		
Phenol, equilibrated	USB	20072
Phenol:chloroform:Isoamyl alcohol	USB	20081
Potassium hydroxide solution, 8.0 N	Sigma	17-8
Potassium chloride	Sigma	P-9541
Sodium chloride	Sigma	S-3014
Sephacryl S-400, high resolution	Pharmacia	17-0609-01
Tubes w/ screw cap, 13 ml	Sarstedt	60.540

The various buffers and other preparations referred to
in the protocol are shown below.

HE buffer is prepared at pH = 7.5 (adjusted with KOH)
by adding 8.34 g of HEPES, free acid (use a better grade than
Sigma's standard; the final concentration is 35 mM), to 200 μl of
0.5 M EDTA, pH 8.0 (final concentration is 0.1 mM) and adding
water to a final volume of 1 L.

HEK buffer is identical to HE buffer but also contains
KCl at a final concentration of 50 mM.

HEKL buffer is identical to HEK buffer but also contains alpha-lactose, which may require warming to go into solution, at a final concentration of 0.2 M.

Lysis buffer (6 ml) is prepared by mixing 4.2 ml of HE buffer with 1ml of 50% glycerol, 750 μl of protease free BSA at 10 mg/ml in PBS, 10 μl of 0.5 M DTT, and 12.5 μl of 0.1M PMSF in isopropanol.

HEK/BSA buffer is prepared by dissolving 5 g of BSA, fraction V, in 500 ml of HEK buffer.

WTEK buffer is prepared at pH = 7.5 by adding 7.53 g of Tris, pH = 7.5 (final concentration of 50 mM), to 20 ml of 0.5 M EDTA (final concentration of 10 mM) and 7.45 g of KCl (final concentration of 100 mM) and adding water to a final volume of 1 L.

TEK buffer is prepared at pH = 7.5 by adding 1.51 g of Tris, pH = 7.5 (final concentration of 10 mM), to 200 μl of 0.5 M EDTA (final concentration of 0.1 mM) and 7.45 g of KCl (final concentration of 100 mM) and adding water to a final volume of 1 L.

The process can be conveniently carried out over a two day period, as shown below.

Day 1

1.   Coat two sets of 12 microtiter wells with the appropriate amount of immobilizing antibody in 100 μl of PBS, for panning and negative control; let the coated plate incubate at 37°C for 1hr. Consider using all 24 wells as "plus receptor" wells in the first round, i.e., no negative control in the first round.

2.   Wash the plate four times (4x) with HEK/BSA.

3.   Block wells by adding 200 μl of HEK/BSA to each well; let the plate incubate at 37°C for 1 hr.

4.   Wash the plate 4x with HEK/BSA.

5.   Dilute the receptor preparation in cold HEK/BSA (or appropriate binding buffer) as necessary.

6.   Add the diluted receptor preparation to the wells at 100 μl per well; let the plate incubate at 4°C for 1 hr. with agitation.

7.   Wash the plate 2x with cold HEK/BSA.

8.   Add 100 μl of 0.1 mg/ml bulk DNA in HEK/BSA to each well; incubate the plate at 4°C for at least 10 minutes.

On day 1, steps A - O should also be carried out.

A.   Begin equilibrating column with cold HEKL (~1hr, flow rate is set to collect 5 ml fractions every 2 to 3 minutes).

B.   Prepare 1 ml of lysozyme at 10 mg/ml in cold HE.

C.   Thaw and combine sub-libraries (2 ml final volume) in a 13 ml Sarstedt screw cap tube.

D.   Add 6 ml of lysis buffer and 150 μl of lysozyme solution (Boehringer lysozyme is preferred over Sigma lysozyme); mix by inverting gently; and incubate on ice for 5 minutes, although less time is often satisfactory.

E.   Add 2 ml of 20% lactose and 250 μl of 2 M KCl, and mix by inverting gently.

F.   Spin at 14.5 K for 15 minutes in a Beckman JA-20 rotor.

G.   Transfer supernatant by pipetting into a new tube.

H.   Load raw lysate onto the equilibrated column.

I.   After lysate is loaded, collect ten 5 ml fractions.

J.   Perform the coomassie protein assay as follows: (1) to 10 microtiter wells, add 100 μl of coomassie reagent and 20 μl from each fraction, and mix; (2) select 4 consecutive fractions which correspond to 1 brown and 3 blue wells from the assay (light blue counts as blue).

K.   Combine selected fractions in a Centriprep100. Two centripreps may be used to speed up the process. The maximum capacity of each centriprep is about 15 ml.

L.   Spin in Beckman J-6B centrifuge at 1500 rpm.

M.   Rinse the column with cold HEK for 1 hr.

N.   Empty liquid from the inner chamber every 15 minutes until final volume < 2 ml (~1 hr.).

O.   Determine lysate volume, and remove 1% as "Pre" sample; keep Pre sample on ice.

Returning to the numbered steps, one proceeds as follows.

9.   Wash plate 2x with cold HEK/BSA.

10.   Bring the volume of the concentrated lysate up to 2400 μl by adding HEKL/BSA; add bulk DNA to a final concentration of

0.1 mg/ml. The activity of the receptor in this buffer should be verified.

11. Add lysate at 100 μl per well; incubate the plate at 4°C for 1 hr. with agitation.

12. Wash plate 4x with cold HEKL/BSA.

13. Add 100 μl of 0.1 mg/ml bulk DNA in HEKL/BSA to each well; incubate at 4°C for 30 minutes with agitation.

14. Wash plate 4x with cold HEKL.

15. Quickly wash plate 1x with cold HEK.

16. Elute by adding to each well 100 μl of a solution composed of 1 mM IPTG, 10 μg/ml of ON-413, and 0.2 M KCl in HE; incubate at room temperature for 30 minutes with agitation. 50 mM KCl may also be used.

17. Remove all eluants; a phenol/CHC$_{13}$ extraction is optional.

18. Add one-tenth volume of 5 M NaCl and 1 μl of 20 mg/ml glycogen as carrier.

19. Precipitate plasmids in equal volume of isopropanol at room temperature.

20. Spin 10 minutes; carefully remove supernatant, spin again, and remove remaining supernatant.

21. Wash with 200 μl of 70% EtOH.

22. Spin and remove traces of supernatant as above.

23. Resuspend plasmids in water (suggested volumes: 100 μl for Pre; and 4 μl each for the panning and negative control wells; use more than 4 μl for panning and negative control samples in later rounds to retain as backups).

Day 2

24. Chill 4 sterile 0.2 cm electrode gap cuvettes on ice. The panning sample is divided equally into 2 cuvettes to prevent complete loss of sample during electroporation.

25. To three 16 ml sterile culture tubes, add 1 ml SOC medium (2% Bacto-Tryptone, 0.5% Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgC$_{12}$, 10 mM MgSO$_4$, and 20 mM Glucose) plus 10 mM NaCitrate to two tubes and 2 ml to one tube. Label the two 1 ml tubes as "Pre" and "NC" (for "negative control"), and label the 2 ml tube as "Pan" (for "panning").

26. Thaw 200 μl of high efficiency electro-competent cells.

27. Transfer 40 μl aliquots of cells to 4 chilled sterile eppendorf tubes; incubate the tubes on ice.

28. Add 2 μl of each plasmid to each Pre and NC tube and 4 μl to the Pan tube and mix gently.

29. Transfer cells/plasmids mixtures into their corresponding cuvettes; keep the cuvettes on ice.

30. Set the BioRad Gene Pulser apparatus to 2.5 kV, 25 μFD capacity, and set the Pulser Controller unit to 200 ohms.

31. Apply one pulse (time constant = 4-5 msec).

32. Immediately add the room temperature SOC-Citrate medium to resuspend cells in the cuvette.

33. Transfer cell suspension back to the culture tube.

34. Incubate the culture tube at 37°C for 1 hr. with agitation.

35. To 200 ml of LB broth, add 2 ml of 1 M sodium citrate and 0.4 ml of 50 mg/ml ampicillin prewarmed to 37°C.

36. Remove 10 to 100 μl of the "Pan" library culture for plating, and transfer the rest (2 ml) to the prewarmed LB broth. Plate out several dilutions of each sample on LB plates containing ampicillin. Suggested plate dilutions are as follows: Pre -- 10^{-5}, 10^{-6} and 10^{-7}; and Pan/NC -- 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6}.

37. Grow "Pan" library at 37°C for about 4-5 hr. until the OD_{600} = 0.5-1.0.

38. Chill the flask rapidly in ice water for at least 10 minutes.

39. Centrifuge cells in 250 ml sterile bottle at 6K for 6 minutes in a Beckman JA-14 rotor.

40. Wash by vortexing cells in 100 ml of cold WTEK.

41. Centrifuge at 6K for 6 minutes.

42. Wash by vortexing cells in 50ml cold TEK.

43. Centrifuge at 6K for 6 minutes.

44. Resuspend cells in 4 ml of HEK and store in two 2 ml vials at -70°C. Use one tube for the next round; keep the other as a backup.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of

clarity of understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

53

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Schatz, Peter J.
                  Cull, Millard G.
                  Miller, Jeff F.
                  Stemmer, Willem P.C.

    (ii) TITLE OF INVENTION: Peptide Library and Screening Method

   (iii) NUMBER OF SEQUENCES: 91

    (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: William M. Smith
        (B) STREET: One Market Plaza, Steuart Tower, Suite 2000
        (C) CITY: San Francisco
        (D) STATE: California
        (E) COUNTRY: USA
        (F) ZIP: 94105

    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

    (vi) CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER: US
        (B) FILING DATE:
        (C) CLASSIFICATION:

   (vii) PRIOR APPLICATION DATA:
        (A) APPLICATION NUMBER: US 07/778,223
        (B) FILING DATE: 16-OCT-1991

  (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: Smith, William M.
        (B) REGISTRATION NUMBER: 30,223
        (C) REFERENCE/DOCKET NUMBER: 11509-50-1

    (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: 415-326-2400
        (B) TELEFAX: 415-326-2422

54

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 14 amino acids
       (B) TYPE: amino acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Gly Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1   5   10

55

(2) INFORMATION FOR SEQ ID NO:2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 54 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGGCGCCNN KNNKNNKNNK NNKNNKNNKN NKNNKNNKNN KNNKTAAGGT CTCG   54

56

(2) INFORMATION FOR SEQ ID NO:3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 11 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TGCCACCGCG G   11

57

(2) INFORMATION FOR SEQ ID NO:4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 14 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATTCCAGAGC TCGA   14

58

(2) INFORMATION FOR SEQ ID NO:5:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: Peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu Glu Ser Gly Gln Gly Ala Asp Gly Ala
1   5   10

59

(2) INFORMATION FOR SEQ ID NO:6:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTCGAGAGCG GGCAGGGGGC CGACGGGGCC TAATTAATTA AGCTT   45

60

(2) INFORMATION FOR SEQ ID NO:7:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 13 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: dynB 1.0

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Tyr Gly Gly Phe Leu Arg Arg Gln Phe Lys Val Val Thr
1   5   10

61

(2) INFORMATION FOR SEQ ID NO:8:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 21 4 1.2

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Thr Gly Lys Arg Gly Phe Lys Val Val Cys Asn Ser
1   5   10

(2) INFORMATION FOR SEQ ID NO:9:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 22 4 1.2

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Lys Arg Asn Phe Lys Val Val Gly Ser Pro Cys Gly
1   5   10

63

(2) INFORMATION FOR SEQ ID NO:10:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 28 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 10 4 0.3

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ser Asp Ser Gly Asn Gly Leu Gly Ile Arg Arg Phe Lys Val Ser Ser
1   5   10   15

Leu Ala Val Leu Ala Asp Glu Arg Arg Phe Ser Ala
            20   25

64

(2) INFORMATION FOR SEQ ID NO:11:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (vii) IMMEDIATE SOURCE:
        (B) CLONE: 30 4 0.9

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Glu Thr Arg Pro Phe Lys Val Ser Glu Tyr Ile Leu
1   5   10

65

(2) INFORMATION FOR SEQ ID NO:12:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 28 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 35 4 0.2

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ser Leu Lys Asp Glu Asn Asn Lys Arg Arg Ile Phe Lys Val Ser Ser
1   5   10   15

Leu Ala Val Leu Ala Asp Glu Arg Arg Phe Ser Ala
            20   25

66

(2) INFORMATION FOR SEQ ID NO:13:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 57 3 0.9

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ser Tyr Leu Arg Arg Glu Phe Lys Val Ser Gly Val
1   5   10

67

(2) INFORMATION FOR SEQ ID NO:14:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 24 4 0.9

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Trp Arg Ser Cys Pro Arg Gln Phe Lys Val Thr
1   5   10

68

(2) INFORMATION FOR SEQ ID NO:15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 45 3 0.9

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Ile Lys Arg Gly Phe Lys Ile Thr Ser Ala Met Ser
1   5   10

(2) INFORMATION FOR SEQ ID NO:16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 47 3 0.8

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Arg Phe Ile Ala Arg Pro Phe Arg Ile Thr Gly
1   5   10

70

(2) INFORMATION FOR SEQ ID NO:17:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 71 2 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ala Arg Ala Phe Arg Val Thr Arg Ile Ala Gly Val
1   5   10

71

(2) INFORMATION FOR SEQ ID NO:18:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 74 2 0.2

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Lys Asn Glu Thr Arg Arg Pro Phe Arg Gln Thr Ala
1   5   10

72

(2) INFORMATION FOR SEQ ID NO:19:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 68 2 0.6

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Val Asn His Arg Arg Phe Ser Val Val His Ser Tyr
1   5   10

73

(2) INFORMATION FOR SEQ ID NO:20:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
     , (B) CLONE: 48 3 0.4

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Val Ser Ser Ser Arg Thr Phe Asn Val Thr Arg Arg
1   5   10

74

(2) INFORMATION FOR SEQ ID NO:21:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 46 3 0.3

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Gly Arg Ser Phe His Val Thr Ser Phe Gly Ser Val
1   5   10

75

(2) INFORMATION FOR SEQ ID NO:22:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 4 4 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Arg Ser Thr Thr Val Arg Gln His Lys Val Val Gly
1   5   10

76

(2) INFORMATION FOR SEQ ID NO:23:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 15 4 1.2

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Glu Arg Pro Asn Arg Leu His Lys Val Val His Ala
1   5   10

(2) INFORMATION FOR SEQ ID NO:24:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 73 2 0.5

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Trp Gln Asn Arg Thr His Lys Val Val Ser Gly Arg
1   5   10

78

(2) INFORMATION FOR SEQ ID NO:25:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 78 2 1:1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Ala Arg Lys His Lys Val Thr
1   5

79

(2) INFORMATION FOR SEQ ID NO:26:

        (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 11 amino acids
                (B) TYPE: amino acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
                (B) CLONE: 40 3 1.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

    Arg Gln Val Thr Arg Leu His Lys Val Ile His
    1   5   10

80

(2) INFORMATION FOR SEQ ID NO:27:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 11 4 1.0

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Cys Pro Gly Glu Arg Met His Lys Ala Val Arg Ala
1   5   10

81

(2) INFORMATION FOR SEQ ID NO:28:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 2 4 1.0

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ser Arg Cys Arg Asn His Arg Val Val Thr Ser Gln
1   5   10

82

(2) INFORMATION FOR SEQ ID NO:29:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 26 4 0.8

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

    Asn Asp Gly Arg Pro His Arg Val Val Arg Cys Gly
    1   5   10

(2) INFORMATION FOR SEQ ID NO:30:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (vii) IMMEDIATE SOURCE:
        (B) CLONE: 9 4 0.8

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Glu Ile Arg Arg His Arg Val Thr Glu Arg Val Asp
1   5   10

84

(2) INFORMATION FOR SEQ ID NO:31:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 56 3 1.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Leu Arg Arg Leu His Arg Val Thr Asn Thr Met Thr
1   5   10

85

(2) INFORMATION FOR SEQ ID NO:32:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 69 2 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Val Lys Gln Arg Leu His Ser Val Val Arg Pro Gly
1   5   10

86

(2) INFORMATION FOR SEQ. ID NO:33:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 7 4 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Val Thr Gln Arg Val Arg Ser Asn Lys Val Val Ser
1   5   10

87

(2) INFORMATION FOR SEQ ID NO:34:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 20 4 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

His Val Glu Lys Ile Lys Arg Leu Asn Lys Val Val
1   5   10

88

(2) INFORMATION FOR SEQ ID NO:35:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 23 4 1.2

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Arg Leu Lys Thr Arg Leu Asn Lys Val Val Met Asp
1   5   10

89

(2) INFORMATION FOR SEQ ID NO:36:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 63 2 0.4

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Val Arg Met Asn Lys Val Val Cys Glu Lys Leu Trp
1   5   10

90

(2) INFORMATION FOR SEQ ID NO:37:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 11 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 49 3 0.3

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Asp Leu Lys Arg Leu Asn Arg Val Val Gly His
1   5   10

91

(2) INFORMATION FOR SEQ ID NO:38:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (vii) IMMEDIATE SOURCE:
        (B) CLONE: 19 4 0.8

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Arg Ile Arg Asn Asn Lys Val Ile Ala Arg Pro Val
1   5   10

92

(2) INFORMATION FOR SEQ ID NO:39:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 12 amino acids
       (B) TYPE: amino acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
       (B) CLONE: 36 4 0.5

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Ser Arg Val Arg Ser Asn Lys Val Ile Met Ser Ile
1   5   10

93

(2) INFORMATION FOR SEQ ID NO:40:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (vii) IMMEDIATE SOURCE:
        (B) CLONE: 77 2 0.6

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Ser Cys Arg Leu Asn Lys Val Ile Ala Arg Pro Val
1   5   10

94

(2) INFORMATION FOR SEQ ID NO:41:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 33 4 0.5

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Arg Ala Leu Ser Lys Asp Arg Leu Asn Lys Val Thr
1   5   10

95

(2) INFORMATION FOR SEQ ID NO:42:

    .(i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 58 3 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Cys Thr Thr Glu Arg Ser Arg Gln Trp Lys Val Thr
1   5   10

96

(2) INFORMATION FOR SEQ ID NO:43:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 16 4 1.1·

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Ala Arg Pro Trp Lys Ile Thr Arg Asn Glu Pro Gly
1   5   ·10

97

        (2) INFORMATION FOR SEQ ID NO:44:

            (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 12 amino acids
                (B) TYPE: amino acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

            (ii) MOLECULE TYPE: peptide

            (vii) IMMEDIATE SOURCE:
                (B) CLONE: 72 2 0.3

            (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

            Gly Val Ser Glu Cys Arg Lys Trp Lys Ile Val Gln
            1   5   10

98

(2) INFORMATION FOR SEQ ID NO:45:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 6 4 1.2

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Thr Thr Leu Arg Arg Tyr Lys Val Thr Gly Glu Arg
1   5   10

99

(2) INFORMATION FOR SEQ ID NO:46:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 34 4 1.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Ile Ala Asp Arg Arg Pro Tyr Arg Val Thr Arg Pro
1   5   10

100

(2) INFORMATION FOR SEQ ID NO:47:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 12 amino acids
          (B) TYPE: amino acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (vii) IMMEDIATE SOURCE:
          (B) CLONE: 76 2 1.2

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

    Ala Gly Lys Val Leu Arg Ala Tyr Lys Ile Val Glu
    1   5   10

101

(2) INFORMATION FOR SEQ ID NO:48:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 12 amino acids
       (B) TYPE: amino acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
       (B) CLONE: 8 4 1.0

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Gln Lys Arg Leu Met Lys Val Ile Phe Glu Gly Arg
1   5   10

102

(2) INFORMATION FOR SEQ ID NO:49:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 55 3 1.0

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Glu Val Pro His Arg Phe Arg Trp Thr Lys His Met
1   5   10

103

(2) INFORMATION FOR SEQ ID NO:50:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 24 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 13 4 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Ser Thr Thr Glu Arg Arg Ser Phe Lys Val Ser Ser Leu Ala Val Leu
1   5   10   15

Ala Asp Glu Arg Arg Phe Ser Ala
            20

104

(2) INFORMATION FOR SEQ ID NO:51:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 23 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 14 4 0.2

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Arg Leu Pro Gly Arg Met Phe Lys Val Ser Ser Leu Ala Val Leu Ala
1   5   10   15

Asp Glu Arg Arg Phe Ser Ala
            20

105

(2) INFORMATION FOR SEQ ID NO:52:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 28 4 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Val Gly Ser Phe Lys Arg Thr Phe Lys Val Ser Cys
1   5   10

106

(2) INFORMATION FOR SEQ ID NO:53:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 21 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 29 4 0.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Arg Gly Arg Met Phe Lys Val Ser Ser Leu Ala Val Leu Ala Asp Glu
1   5   10   15

Arg Arg Phe Ser Ala
            20

(2) INFORMATION FOR SEQ ID NO:54:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 29 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 54 3 0.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Pro Gly Arg Trp Val Arg Gly Val Gly Ile Arg Cys Phe Lys Val Ser
1   5   10   15

Ser Leu Ala Val Leu Ala Asp Glu Arg Arg Phe Ser Ala
        20   25

108

(2) INFORMATION FOR SEQ ID NO:55:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 22 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (vii) IMMEDIATE SOURCE:
        (B) CLONE: 60 2 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Arg Met Ser Arg Leu Phe Lys Val Ser Ser Leu Ala Val Leu Ala Asp
1   5   10   15

Glu Arg Arg Phe Ser Ala
            20

109

(2) INFORMATION FOR SEQ ID NO:56:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 28 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 1 4 0.1

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Pro Asp Val Leu Arg Ala Val Ala Thr Arg Gln His Lys Val Ser Ser
1   5   10   15

Leu Ala Val Leu Ala Asp Glu Arg Arg Phe Ser Ala
            20   25

110

(2) INFORMATION FOR SEQ ID NO:57:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

 (vii) IMMEDIATE SOURCE:
        (B) CLONE: 27 4 0.2

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Arg Val Arg Gly His Arg Val Val Met Tyr Asn Glu
1   5   10

111

(2) INFORMATION FOR SEQ ID NO:58:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 64 2 0.1

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Glu Cys Leu His Arg Arg Val His Lys Ile Leu Ser
1   5   10

112

(2) INFORMATION FOR SEQ ID NO:59:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 61 2 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

    Gly Leu Lys Cys Arg Pro Met Lys Val Asn Ala Asp
    1   5   10

113

(2) INFORMATION FOR SEQ ID NO:60:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 50 3 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Arg His Arg Pro Phe Gly Trp Val Asn Lys Arg Ser
1   5   10

114

(2) INFORMATION FOR SEQ ID NO:61:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (vii) IMMEDIATE SOURCE:
        (B) CLONE: 52 3 0.2

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Ala Arg Leu Phe Ser Gln Ile Arg Arg Phe Pro
1   5   10

115

(2) INFORMATION FOR SEQ ID NO:62:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 53 3 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Arg Val Arg Trp His Met Val Thr Gly Asp Lys Gly
1   5   10

116

(2) INFORMATION FOR SEQ ID.NO:63:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 31 4 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Arg Phe Arg Asn Cys Ser Ile Ile Ser Ala Arg Gly
1   5   10

117

(2) INFORMATION FOR SEQ ID NO:64:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (vii) IMMEDIATE SOURCE:
        (B) CLONE: 62 2 0.1

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

    Tyr Gly Val Pro Arg Ile val Ala His Gln Leu Met
    1   5   10

118

(2) INFORMATION FOR SEQ ID NO:65:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 5 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Gly Ala Asp Gly Ala
1   5

119

(2) INFORMATION FOR SEQ ID NO:66:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 13 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Tyr Gly Gly Phe Leu Arg Arg Gln Phe Lys Val Val Thr
1   5   10

120

(2) INFORMATION FOR SEQ ID NO:67:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

   Gly Lys Arg Xaa
   1

121

(2) INFORMATION FOR SEQ ID NO:68:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 53 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GCGGGCTAGC TAACTAATGG AGGATACATA AATGAAACCA GTAACGTTAT ACG   53

122

(2) INFORMATION FOR SEQ ID NO:69:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 46 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGTTCCGAGC TCACTGCCCG CTCTCGAGTC GGGAAACCTG TCGTGC   46

123

(2) INFORMATION FOR SEQ ID NO:70:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 51 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CCTCCATATG AATTGTGAGC GCTCACAATT CGGTACAGCC CCATCCCACC C   51

124

(2) INFORMATION FOR SEQ ID NO:71:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 51 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CAGCATCGAT CAATTGTGAG CGCTCACAAT TCAGGATGTG TGTGATGAAG A   51

125

(2) INFORMATION FOR SEQ ID NO:72:

        (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 72 base pairs
                (B) TYPE: nucleic acid
                (C) STRANDEDNESS: single
                (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TCGAGAGCGG GCAGGGGGCC GACGGGGCCT ACGGTGGTTT CCTGCGTCGT CAGTTCAAAG   60

TTGTAACCTA AT   72

126

(2) INFORMATION FOR SEQ ID NO:73:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 72 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CTAGATTAGG TTACAACTTT GAACTGACGA CGCAGGAAAC CACCGTAGGC CCCGTCGGCC   60

CCCTGCCCGC TC   72

127

(2) INFORMATION FOR SEQ ID NO:74:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 15 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GGGCCTAATT AATTA   15

128

(2) INFORMATION FOR SEQ ID NO:75:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 22 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

AGCTTAATTA ATTAGGCCCC GT   22

129

(2) INFORMATION FOR SEQ ID NO:76:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 54 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTGGCGCCNN KNNKNNKNNK NNKNNKNNKN NKNNKNNKNN KNNKTAAGGT CTCG   54

130

(2) INFORMATION FOR SEQ ID NO:77:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 11 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGCGCCACCG T   11

131

(2) INFORMATION FOR SEQ ID NO:78:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 14 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

AGCTCGAGAC CTTA   14

132

(2) INFORMATION FOR SEQ ID NO:79:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 9 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Asp Tyr Met Gly Trp Met Asp Phe Gly
1   5

133

(2) INFORMATION FOR SEQ ID NO:80:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Arg Gln Phe Lys Val Val
1   5

134

(2) INFORMATION FOR SEQ ID NO:81:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Arg Gln Phe Lys Val Val Thr
1   5

135

(2) INFORMATION FOR SEQ ID NO:82:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 21 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

TATTTGCACG GCGTCACACT T   21

136

(2) INFORMATION FOR SEQ ID NO:83:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 47 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

CCGCGCCTGG GCCCAGGGAA TGTAATTGAG CTCCGCCATC GCCGCTT   47

137

(2) INFORMATION FOR SEQ ID NO:84:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 62 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

CGATGGCGGA GCTCAATTAC ATTCCCNNKN NKNNKNNKNN KAAACCAGTA ACGTTATACG   60

AT   62

138

(2) INFORMATION FOR SEQ ID NO:85:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 59 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

CGATGGCGGA GCTCAATTAC ATTCCCNNKN NKNNKNNKAA ACCAGTAACG TTATACGAT   59

139

(2) INFORMATION FOR SEQ ID NO:86:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 72 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

CGCCCGCCAA GCTTAGGTTA CAACTTTGAA CTGACGMNNM NNMNNMNNGG GAATGTAATT   60

CAGCTCCGCC AT   72

140

(2) INFORMATION FOR SEQ ID NO:87:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 32 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GAATTCAATT GTGAGCGCTC ACAATTGAAT TC   32

141

(2) INFORMATION FOR SEQ ID NO:88:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 93 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: 1..85

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

```
CTC GAG AGC GGG CAG GTG GTG CAT GGG GAG CAG GTG GGT GGT GAG GCC          48
Leu Glu Ser Gly Gln Val Val His Gly Glu Gln Val Gly Gly Glu Ala
 1           5               10              15

TCC GGG GCC GTT AAC GGC CGT GGC CTA GCT GGC CAA T AAGTCGAC               93
Ser Gly Ala Val Asn Gly Arg Gly Leu Ala Gly Gln
             20              25
```

142

(2) INFORMATION FOR SEQ ID NO:89:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 28 amino acids
            (B) TYPE: amino acid
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: protein

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Leu Glu Ser Gly Gln Val Val His Gly Glu Gln Val Gly Gly Glu Ala
 1   5   10   15

Ser Gly Ala Val Asn Gly Arg Gly Leu Ala Gly Gln
            20   25

143

(2) INFORMATION FOR SEQ ID NO:90:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 93 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: 1..60

    (ix) FEATURE:
        (A) NAME/KEY: misc_feature
        (B) LOCATION: 61..63
        (D) OTHER INFORMATION: /note: "NNK can appear up to n times
                              where n=5,6,7,8 or more"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

```
CTC GAG AGC GGG CAG GTG GTG CAT GGG GAG CAG GTG GGT GGT GAG GCC        48
Leu Glu Ser Gly Gln Val Val His Gly Glu Gln Val Gly Gly Glu Ala
 1           5                  10                 15

TCC GGA GGT GGT NNK TAACTAAGTA AAGCTGGCCA ATAAGTCGAC                    93

Ser Gly Gly Gly Xaa
            20
```

144

(2) INFORMATION FOR SEQ ID NO:91:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 21 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Leu Glu Ser Gly Gln Val Val His Gly Glu Gln Val Gly Gly Glu Ala
1   5   10   15

Ser Gly Gly Gly Xaa
            20

WHAT IS CLAIMED IS:

1.   A method of constructing a random peptide library of at least 10^6 members, said method comprising the steps of:

    (a)   constructing a recombinant DNA vector that encodes a DNA binding protein and contains a binding site for the DNA binding protein;

    (b)   inserting into the coding sequence of the DNA binding protein in at least 10^6 vectors of step (a) a coding sequence for a random peptide such that the resulting vectors encode at least 10^6 different fusion proteins, each of which is composed of the DNA binding protein and a random peptide;

    (c)   transforming host cells with the vectors of step (b); and

    (d)   culturing the host cells transformed in step (c) under conditions suitable for expression of the fusion proteins.

2.   The method of Claim 1, wherein said host cell is a bacterium.

3.   The method of Claim 2, wherein said bacterium is E. coli, and said recombinant DNA vector is a plasmid.

4.   The method of Claim 3, wherein said DNA binding protein is selected from the group of proteins consisting of phage repressor or activator proteins, transcriptional regulators, phage 434 repressor, lambda phage cI and cro repressors, E. coli CAP protein, myc and related proteins, fos protein, jun protein, Drosophila paired protein, TFIIIA, yeast Gal4, phage P22 Arc and Mnt repressors, lac repressor, and protein complexes comprising either yeast Gal80 or adenovirus E1A protein.

5.   The method of Claim 4, wherein said DNA binding protein is a lac repressor protein composed of two lac headpieces joined by a linker.

6.   The method of Claim 4, wherein said DNA binding protein is the lac repressor protein, said DNA binding site is either lacO or $lacO_s$, and said plasmid contains at least two DNA binding sites.

7.   The method of Claim 6, wherein said random peptide is located at the carboxy terminus of said fusion protein.

8.   A method for screening a random peptide library of Claim 1, said method comprising the steps of:

(a)   lysing the cells transformed with the peptide library under conditions such that the fusion protein remains bound to the vector that encodes the fusion protein;

(b)   contacting the fusion proteins of the peptide library with a receptor under conditions conducive to specific peptide - receptor binding; and

(c)   isolating the vector that encodes a peptide that binds to said receptor.

9.   The method of Claim 8 further comprising the steps of:

(d)   transforming a host cell with the vectors obtained in step (c); and repeating steps (a), (b), and (c) with the host cells transformed in step (d).

10.   The method of Claim 9, wherein said host cell is E. coli.

11.   The method of Claim 10, wherein said DNA binding protein is a lac repressor protein and said DNA binding site is either lacO or lacO$_s$.

12.   The method of Claim 11, wherein said vector is a plasmid that contains at least two lacO$_s$ DNA binding sites.

13.   A recombinant DNA vector useful for constructing a random peptide library, said vector comprising:

(a)   a DNA sequence encoding a DNA binding protein;

(b)   a promoter positioned so as to drive transcription of said DNA binding protein coding sequence;

(c)   at least two binding sites for said DNA binding protein; and

(d)   a coding sequence for a peptide inserted in said DNA binding protein coding sequence so that said coding sequences can be transcribed to produce an RNA transcript that can be translated to produce a fusion protein capable of binding to said DNA binding sites.

14.   The plasmid of Claim 13, wherein said DNA binding protein is a lac repressor protein and said DNA binding sites are either lacO or lacO$_s$.

15.   The vector of Claim 14 that is plasmid pMC3.

16.   The vector of Claim 14 that is plasmid pMC5.

17.   The vector of Claim 14 that is plasmid pJS123.

18.   The vector of Claim 14 that is plasmid pJS141.

19.   The vector of Claim 14 that is plasmid pJS142.

5   20.   A recombinant host cell transformed with a vector
of Claim 13.

21.   The transformed host cell of Claim 20 that is
E.coli ARI 161/pMC3.

22.   The transformed host cell of Claim 20 that is E.
10   coli ARI 161/pMC5.

23.   The transformed host cell of Claim 20 that is E.
coli ARI 161/pJS123.

24.   The transformed host cell of Claim 20 that is E.
coli ARI 246/pJS141.

15   25.   The transformed host cell of Claim 20 that is E.
coli ARI 280/pJS142.

26.   A random peptide library composed of at least 10^6
different members, wherein each member is a host cell transformed
with a recombinant DNA vector that encodes a DNA binding protein
20   and contains a binding site for the DNA binding protein and a
coding sequence for a random peptide inserted into the coding
sequence of the DNA binding protein such that the resulting
vector encodes a fusion protein that is composed of the DNA
binding protein and the random peptide; and wherein each
25   different member differs from other members with respect to the
sequence of the random peptide.

27.   A ligand fragment library composed of at least 10
different members, wherein each member is a host cell transformed
with a recombinant DNA vector that encodes a DNA binding protein
30   and contains a binding site for the DNA binding protein and a
coding sequence for a ligand fragment inserted into the coding
sequence of the DNA binding protein such that the resulting
vector encodes a fusion protein that is composed of the DNA
binding protein and the ligand fragment; and wherein each
35   different member differs from other members with respect to the
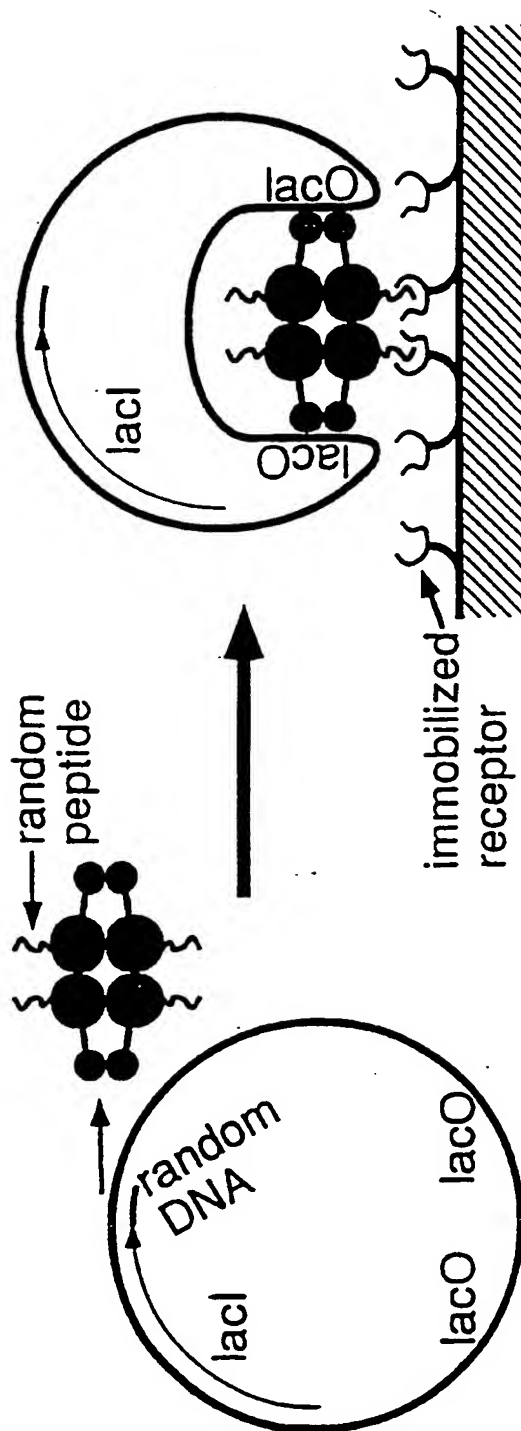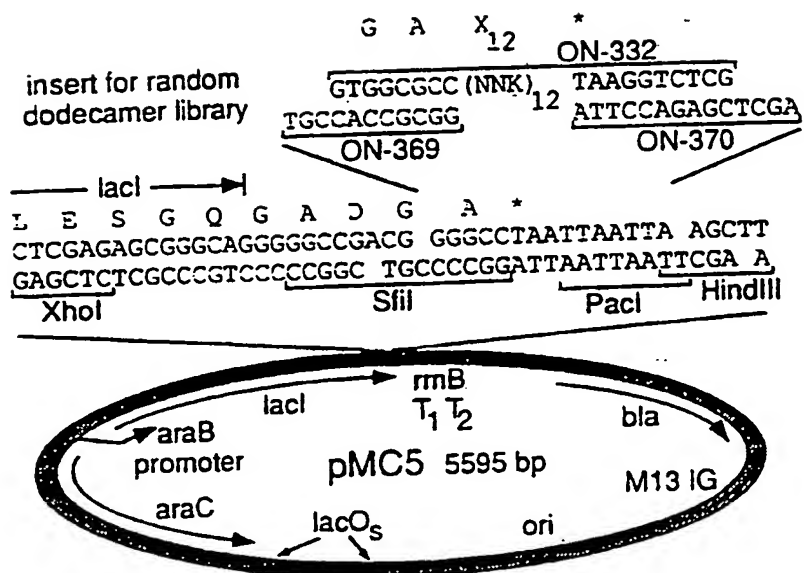sequence of the ligand fragment.

FIGURE 1

FIGURE 2

#/Rnd/ELISA			Peptide Sequence		
dynB		1.0	. . . . Y G G F L R	R Q F K V V	T
21	4	1.2	. . . . T G K	R G F K V V	C N S
22	4	1.2	. . . . K	R N F K V V	G S P C G
10	4	0.3	. . S D S G N G L G I	R R F K V S	S *
30	4	0.9	. . . . G T	R P F K V S	E Y I L
35	4	0.2	. . S L K D E N N K R	R I F K V S	S *
57	3	0.9	. . . . S Y L R	R E F K V S	G V
24	4	0.9	. . . . . . G W R S C P	R Q F K V T	
45	3	0.9	. . . . I K	R G F K I T	S A M S
47	3	0.8	. . . . . . V R F I A	R P F R I T	G
71	2	1.1	. . . . A	R A F R V T	R I A G V
74	2	0.2	. . . . . . K N E T R	R P F R Q T	A
68	2	0.6	. . . . V N H	R R F S V V	H S Y
48	3	0.4	. . . . V S S S	R T F N V T	R R
46	3	0.3	. . . . G	R S F H V T	S F G S V
4	4	1.1	. . . . . . R S T T V	R Q H K V V	G
15	4	1.2	. . . . . . E R P N	R L H K V V	H A
73	2	0.5	. . . . W Q N	R T H K V V	S G R
78	2	1.1	. . . . A	R K H K V T	
40	2	1.1	. . . . . . R Q V T	R L H K V I	H
11	4	1.0	. . . . . C P G E	R M H K A V	R A
2	4	1.0	. . . . S R C	R N H R V V	T S Q
26	4	0.8	. . . . N D G	R P H R V V	R C G
9	4	0.8	. . . . E I	R R H R V T	E R V D
56	3	1.1	. . . . L R	R L H R V T	N T M T
69	2	1.1	. . . . V K Q	R L H S V V	R P G
7	4	1.1	. . . . . V T Q R V	R S N K V V	S
20	4	1.1	. . . . H V E K I K	R L N K V V	
23	4	1.2	. . . . . . R L K T	R L N K V V	M D
63	2	0.4	 V	R M N K V V	C E K L W
49	3	0.3	. . . . . . D L K	R L N R V V	G H
19	4	0.8	. . . . R I	R N N K V I	A Y H S
36	4	0.5	. . . . . . S R V	R S N K V I	M S I
77	2	0.6	. . . . S C	R L N K V I	A R P V
33	4	0.5	. . . . . R A L S K D	R L N K V T	
58	3	1.1	. . . . . C T T E R S	R Q W K V T	
16	4	1.1	. . . . A	R P W K I T	R N E P G
72	2	0.3	. . . . . G V S E C	R K W K I V	Q
6	4	1.2	. . . . T T L	R R Y K V T	G E R
34	4	1.1	. . . . . . I A D R	R P Y R V T	R P
76	2	1.2	. . . . . . A G K V L	R A Y K I V	E
8	4	1.0	. . . . Q K	R L M K V I	F E G R
55	3	1.0	. . . . . . E V P H R F	R W T K H M	
13	4	0.1	. . . . S T T G R	R S F K V S	S *
14	4	0.2	. . . . R L P G	R M F K V S	S *
28	4	0.1	. . . . . . V G S F K	R T F K V S	C
29	4	0.1	. . . . R G	R M F K V S	S *
54	3	0.1	. P G R W V R G V G I	R C F K V S	S *
60	2	0.1	. . . . R M S	R L F K V S	S *
1	4	0.1	. . P D V L R A V A T	R Q H K V S	S *
27	4	0.2	. . . . R V	R G H R V V	M Y N E
64	2	0.1	. . . . . . E C L H R	R V H K I L	S
61	2	0.1	. . . . . . G L K C	R P M K V N	A D
50	3	0.1	. . . . R H	R P F G W V	N K R S
52	3	0.2	. . . . A A	R L F S Q I	R R F P
53	3	0.1	. . . . R V	R W H M V T	G D K G
31	4	0.1	. . . . R F	R N C S I I	S A R G
62	2	0.1	. . . . Y G V P	R I V A H Q	L M

ELISA⁺ (rows dynB through 55)

ELISA⁻ (rows 13 through 62)

* = L A V L A D E R R F S A

FIGURE 3

pJS141 (cys free lacI) and pJS142 (WT lacI) Library Vectors: (SEQ. ID NOS. 88, 89)

```
                                                                                Eag I
                                        Sfl I                    Sfl I  Msc I    SalI
                                Stu I         Hpa I
        ---lacI------><---------------linker-------------->
Xho I
L  E  S  G  Q  V  V  H  G  E  Q  V  G  G  E  A  S  G  A  V  N  G  R  G  L  A  G  Q              99
CTCGAGAGCGGGCAGgtggtgcatggggagcaggtgggtggtgagGCCTCCGGGGCCGTTAACGGGCCGTGGCCTAGCTGGCCAA TAA gtcgac
GAGCTCTCGCCCGTCcaccacgtaccccctcgtccaccactcCGGAGGCCCCGGCAATTGCCGGCACCGGATCGACCGGTTATTcagctg
```

Library Construction after SfiI digestion: (SEQ. ID NOS. 90, 91)

```
                                        Stu I                                          Msc I    SalI
                                                                BspEI
        ---lacI------><---------------linker-------------><library>
Xho I
L  E  S  G  Q  V  V  H  G  E  Q  V  G  G  E  A  S  G  G  G     Xn         *
CTCGAGAGCGGGCAGgtggtgcatggggagcaggtgggtggtgagGCCTCCG gaggtggt (NNK)n taactaagt aaagc TGGCCA TAA gtcgac
GAGCTCTCGCCCGTCcaccacgtaccccctcgtccaccactcCGGA ggcctccacca                    attgattcatt TCGACCGGTTATTcagctg
```

FIGURE 4

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12N 15/10, 15/63, 1/21
US CL :435/172.3, 320.1, 252.33

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/172.3, 320.1, 252.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CHEMICAL ABSTRACTS

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 4,643,969 (Inouye et al.) 17 February 1987, see Figure 29.	1-27
Y	Science, Volume 249, issued 27 July 1990, J. K. Scott et al., "Searching for Peptide Ligands with an Epitope Library", pages 386-390, especially the Abstract, Figure 1, and page 389.	1-27
Y	Angew. Chem. Int. Ed. Engl., Volume 30, No. 3, issued 1991, A. Plückthun et al., "The Rationality of Random Screening—Efficient Methods of Selection of Peptides and Oligonucleotide Ligands", pages 296-298, entire document.	1-27
Y	Nature, Volume 249, issued 07 June 1974, B. Müller-Hill et al., "Lac repressor can be fused to ß-galactosidase", pages 561-563, entire document.	1-27
Y	Proceedings of the National Academy of Sciences (USA), Volume 80, issued November 1983, J. R. Sadler et al., "A perfectly symmetric lac operator binds the lac repressor very tightly", pages 6785-6789, especially the Abstract.	1-27

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 JANUARY 1993	26 JAN 1993

Name and mailing address of the ISA/	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JAMES KETTER
Facsimile No. NOT APPLICABLE	Telephone No. (703) 308-1169

Form PCT/ISA/210 (second sheet)(July 1992)*